



PCT

WORLD INTELLECT
Inte

INTERNATIONAL APPLICATION PUBLISHED

(51) International Patent Classification⁶:C12N 15/12, C07K 14/47, 14/705, G01N
33/566, A61K 38/17, 47/48, 51/00, C07K
16/18 // C12N 15/62, C07K 19/00A²

(43) International Publication Date:

8 February 1996 (08.02.96)

(21) International Application Number:

PCT/IB95/00627

(22) International Filing Date:

21 July 1995 (21.07.95)

(30) Priority Data:

110464

26 July 1994 (26.07.94)

IL

(71) Applicant (for all designated States except US): RAMOT
UNIVERSITY AUTHORITY FOR APPLIED RESEARCH
AND INDUSTRIAL DEVELOPMENT LTD. [IL/IL]; 32
Haim Levanon Street, 64731 Tel Aviv (IL).

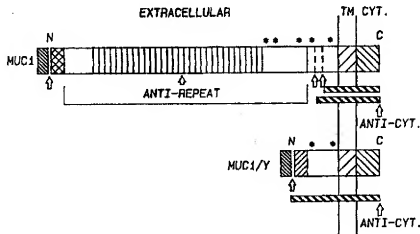
(72) Inventor; and

(75) Inventor/Applicant (for US only): WRESCHNER, Daniel, H.
[GB/IL]; Mevo Hadas 14, P.O. Box 1129, 90435 Efrat (IL).(74) Agent: AUSTIN, Hedley, William; Urquhart-Dykes & Lord,
Alexandra House, 1 Alexandra Road, Swansea SA1 5ED
(GB).(81) Designated States: AM, AU, BB, BG, BR, BY, CA, CN, CZ,
EE, GE, HU, JP, KG, KP, KR, KZ, LK, LR, LT, LV, MD,
MG, MN, MX, NO, NZ, PL, RO, RU, SD, SI, SK, TJ, TT,
UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK,
ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI
patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE,
SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published

Without international search report and to be republished
upon receipt of that report.

(54) Title: MUCIN-DERIVED PROTEINS FOR THE DIAGNOSIS, IMAGING, AND THERAPY OF HUMAN CANCER



(57) Abstract

The invention provides a biochemically pure MUC1 protein, selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, MUC1/W, MUC1/W/alt, MUC1/Z and MUC1/Z/alt or a functional derivative thereof devoid of a tandem repeat array. The invention also provides a diagnostic agent for the detection of human cancer cells and a pharmaceutical composition for the treatment of human cancer, and a method for detecting the presence of cancer in a patient.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

- 1 -

MUCIN-DERIVED PROTEINS FOR THE DIAGNOSIS, IMAGING, AND THERAPY OF HUMAN CANCER**TECHNICAL FIELD**

The present invention relates to a newly-discovered group of protein products of the MUC1 gene and diagnostic and therapeutic methods for utilizing the same, as well as diagnostic and therapeutic compositions containing the same.

BACKGROUND OF THE INVENTION

Polymorphic, high molecular weight glycoproteins are abundantly expressed in human breast carcinomas. These proteins, designated MUC1 (also referred to as episialin, H23Ag, PEM, EMA, CA15-3, MCA, etc.) are heavily glycosylated with O-glycosidic-linked carbohydrate side chains, and, as such, have mucin-like characteristics [for review, see J. Hilkens, et al., "Cell Membrane-Associated Mucins and Their Adhesion Modulating Property," TIBS, Vol. 17, pp. 359-363 (1992)]. Although MUC1 proteins are expressed at basal levels by most secretory epithelial tissues, their expression is dramatically increased in malignant breast epithelial cells [P.X. Xing, et al., "Reactivity of Anti-Human Milk Fat Globule Antibodies with Synthetic Peptides," J. Immunol., Vol. 142, pp. 3503-3509 (1989)]. The fact that disease status in breast cancer patients is routinely assessed by monitoring the serum levels of circulating tandem repeat array containing MUC1 protein, using commercial assays such as CA15-3 and MCA (mammary carcinoma antigen) underscores the unequivocal importance of MUC1 gene expression to human breast cancer. That increased MUC1 expression may reflect a change in the differentiation

- 2 -

status of the malignant epithelial cells is indicated by high levels of MUC1 expression also in lactating mammary epithelial tissue, where it is localized at the apical surfaces. Due to the loss of cellular architecture in breast cancer tissue, MUC1 is no longer expressed solely on the apical surface and this, in conjunction with the finding that MUC1 expression reduces cell-cell adhesion [M.J.L. Ligtenberg, et al., "Suppression of Cellular Aggregation by High Levels of Episialin," Cancer Res., Vol. 52, pp. 2318-2324 (1992)], may enhance the invasiveness of the breast cancer cell.

Molecular studies, including cDNA and gene cloning, have elucidated many properties of the MUC1 proteins [D.H. Wreschner, et al., "Isolation and Characterization of Full Length cDNA Coding for the H23 Breast Tumor Associated Antigen," in Breast Cancer: Progress in Biology, Clinical Management and Prevention, M.A. Rich, J.C. Hager and I. Keydar, Eds., Kluwer Academic Publishers, Boston, Mass., U.S.A., pp. 41-59 (1989); D.H. Wreschner, et al., "Human Epithelial Tumor Antigen cDNA Sequences - Differential Splicing May Generate Multiple Protein Forms," Eur. J. Biochem., Vol. 189, pp. 463-473 (1990)]. The MUC1 gene product best characterized so far is a polymorphic, type 1 transmembrane molecule that consists of a large extracellular domain, a transmembrane domain and a 69 amino acid cytoplasmic tail. The genetic polymorphism derives from a 20 amino acid repeat motif rich in serine, threonine and proline residues, that varies in number from approximately 20 to 100 repeats. The feature of a tandemly repeating domain is shared by all cloned human, porcine and *Xenopus* mucins (MUC2, MUC3, human tracheobronchial mucin MUC4, MUC5, porcine submaxillary mucin and *Xenopus* integumentary mucin).

- 3 -

This common property notwithstanding, several unique features distinguish the MUC1 proteins from the other mucins. First, whereas the latter mucins have several cysteine residues in their extracellular domains that form disulfide bridges, thereby generating a mucin network, the MUC1 proteins have no cysteine residues in their extracellular domain, and thus are less likely to have this mesh-forming capability. Second, and perhaps most significantly, the MUC1 protein is a type 1 transmembrane protein, a molecular structure not shared by the other mucin molecules, that are secreted from the cell.

Insights into the function of MUC1 gene products have been furnished by analyzing the phenotype of tandem repeat array containing transmembrane MUC1 transfectants. This has shown that MUC1 expression reduces cellular adhesion [Litgenberg, et al., Cancer Res., *ibid.* (1992)]. Interestingly, a comparison of the human MUC1 amino acid sequence with the mouse MUC1 homologue [A.P. Spicer, et al., "Molecular Cloning of the Mouse Homologue of the Tumor Associated Mucin, MUC1, Reveals Conservation of Potential O-Glycosylation Sites, Transmembrane and Cytoplasmic Domains and a Loss of Minisatellite-Like Polymorphism," J. Biol. Chem., Vol. 266, pp. 15099-15109 (1991)] shows that whereas a tandem repeat structure rich in serine and threonine residues is also observed in the mouse protein, there is very little conservation of actual amino acid sequence in this region. This indicates that perhaps the primary function of mucin tandemly repeated domains is to provide the "infrastructure" for extensive O-linked glycosylation, thereby conferring to the molecule its anti-adhesion function. Recent experiments have indeed shown that the tandem repeat array mediates this anti-adhesive feature of MUC1 protein.

- 4 -

As described above, expression of the polymorphic MUC1 proteins reduces cellular aggregation potential, suggesting that MUC1 interference with cellular interactions may be critical in tissue morphogenesis such as ductal development by glandular epithelial cells in normal tissues [J. Hilken, et al., *ibid.*, (1992)], and could be responsible for the detachment of tumor cells from malignant tissues where it is expressed at high levels [Ligtenberg, et al., Cancer Res., *ibid.* (1992)].

Comparison of MUC1 sequences in different species may provide additional insights into functionally important regions of MUC1 gene products. For example, the mouse MUC1 homologue shows, in contrast to the lack of similarity within the tandem repeating sequence, a very high degree of amino acid sequence conservation with human MUC1, in the cytoplasmic and transmembrane domains as well as in the 120 amino acids N-terminal to the transmembrane domain. This degree of amino acid sequence similarity is almost 90% in the cytoplasmic and transmembrane domains, indicating that these regions, as well as the 120 amino acids N-terminally adjacent to the transmembrane domain, may be functionally very important. This contrasts with the lack of inter-species conservation of the MUC1 tandem repeat array amino acid sequence, thereby suggesting that distinct functions may be performed by the tandem repeat array and by the other highly-conserved regions of the MUC1 proteins.

SUMMARY OF THE INVENTION

According to the present invention, there has now been identified and characterized a group of novel protein products of the MUC1 gene.

- 5 -

More particularly, the present invention relates to novel proteins designated herein as MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, MUC1/W, MUC1/W/alt, MUC1/Z and MUC1/Z/alt, which function as receptor proteins and activating ligands for said receptors in human breast cancer cells, and which proteins are all characterized by the absence of the characteristic MUC1 protein tandem repeat array.

Thus, according to the present invention, there is now provided a biochemically pure MUC1 protein, selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, MUC1/W, MUC1/W/alt, MUC1/Z, and MUC1/Z/alt, or a functional derivative thereof, devoid of a tandem repeat array.

The term "functional derivative" as used herein is intended to include labelled proteins, conjugated proteins, fused chimeric proteins and purified receptors in soluble form, as well as fragments, deletions, and conservative substitutions of said proteins.

As will be realized, the biochemically pure MUC1 proteins as defined and claimed herein are isolated and purified and are thus substantially free of natural contaminants.

The term "conservative substitutions" as used herein is intended to denote substitutions which preserve the activity of the defined proteins, involving between 80% to 90% conservation.

More specifically, the present invention provides a biochemically pure MUC1 protein selected from the group

- 6 -

consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/W, MUC1/W/alt, MUC1/Z, and MUC1/Z/alt, or a functional derivative thereof, comprising a partial amino acid sequence:

```
M T P G T Q S P F F L L L L L T V L T [A T T A P K P A T]  
V V T G S G H A S S T P G G E K E T S A T Q R S S V P  
S S T E K N A
```

and devoid of a tandem repeat array downstream thereof.

Especially, the present invention provides a biochemically pure MUC1 protein selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/W, MUC1/W/alt, MUC1/Z, and MUC1/Z/alt, or a functional derivative thereof, having a partial amino acid sequence:

```
M T P G T Q S P F F L L L L L T V L T [A T T A P K P A T]  
V V T G S G H A S S T P G G E K E T S A T Q R S S V P  
S S T E K N A
```

and devoid of a tandem repeat array downstream thereof.

Furthermore, the present invention provides a biochemically pure MUC1 protein selected from the group consisting of MUC1/V, MUC1/V/alt, or a functional derivative thereof, comprising a partial amino acid sequence:

```
M T P G T Q S P F F L L L L L T V L T [A T T A P K P A T]  
V V T G S G H A S S T P G G E K E T S A T Q R S S V P S
```

and devoid of a tandem repeat array downstream thereof.

- 7 -

Still furthermore, the present invention provides a biochemically pure MUC1 protein selected from the group consisting of MUC1/V, MUC1/V/alt, or a functional derivative thereof, having a partial amino acid sequence:

M T P G T Q S P F F L L L L L T V L T [A T T A P K P A T]
V V T G S G H A S S T P G G E K E T S A T Q R S S V P S

and devoid of a tandem repeat array downstream thereof.

The sequence starts at the amino (NH₂) terminal methionine (M) residue. The 9 amino acid sequence presented in brackets [A T T A P K P A T] represents an isoform that is generated by an alternative splice acceptor site. Hereinafter, MUC1 derivatives containing this additional 9 amino acid sequence will be referred to as the "/alt configuration" of the novel MUC1 derivatives described herein. The two arrows indicate the sites at which cleavage of the signal sequence is expected to occur (Fig. 2).

Specifically, the present invention provides biochemically pure MUC1/X and MUC1/X/alt, respectively comprising the sequences shown in Figs. 5A and 5B and functional derivatives thereof; biochemically pure MUC1/Y and MUC1/Y/alt respectively comprising the sequences shown in Figs. 6A and 6B and functional derivatives thereof; biochemically pure MUC 1/V, MUC1/V/alt, respectively comprising the sequences shown in Figs. 6C and 6D and functional derivatives thereof; MUC1/W and MUC1/W/alt respectively comprising the sequences shown in Figs. 7A and 7B and functional derivatives thereof; and biochemically pure MUC1/Z and MUC1/Z/alt respectively comprising the

- 8 -

sequences shown in Figs. 8A and 8B and functional derivatives thereof.

More particularly, the present invention provides biochemically pure MUC1/X and MUC1/X/alt, respectively having the sequences shown in Figs. 5A and 5B and functional derivatives thereof; biochemically pure MUC1/Y and MUC1/Y/alt respectively having the sequences shown in Figs. 6A and 6B and functional derivatives thereof; biochemically pure MUC 1/V, MUC1/V/alt, respectively comprising the sequences shown in Figs. 6C and 6D and functional derivatives thereof; MUC1/W and MUC1/W/alt biochemically pure MUC1/W and MUC1/W/alt respectively having the sequences shown in Figs. 7A and 7B and functional derivatives thereof; and biochemically pure MUC1/Z and MUC1/Z/alt respectively having the sequences shown in Figs. 8A and 8B and functional derivatives thereof.

MUC1/X and MUC1/Y have been found to be generated by a splicing mechanism, using perfect splice donor and splice acceptor sites, located upstream and downstream to the tandem repeat array of MUC1 while maintaining the original reading frame, and therefore these proteins retain the cytoplasmic and transmembrane domains, as well as the amino acids immediately N-terminal to the transmembrane domain (Figs. 1A and 1B, Fig. 2, Fig. 3 and Fig. 4).

MUC1/V has been found to be generated by a splicing mechanism, using a different splice donor and splice acceptor sites, located upstream and downstream to the tandem repeat array of MUC1 while also maintaining the original reading frame and therefore these proteins retain the cytoplasmic and transmembrane domains, as well as the

- 9 -

amino acids immediately N-terminal to the transmembrane domain.

On the other hand, MUC1/W and MUC1/Z are generated by a splicing mechanism in which the original reading frame is not maintained and therefore the proteins do not include the cytoplasmic and transmembrane domains (Figs. 1A and 1B, Fig. 2, Fig. 3 and Fig. 4) and are therefore secreted from the cell.

Further extensive research, testing and analysis indicate that MUC1/X, MUC1/Y, MUC1/V and their /alt configurations serve as receptor proteins in breast cancer cells, while MUC1/W and MUC1/Z and their /alt configurations function as ligands for said receptors.

In contrast to the new MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, and MUC1/V/alt proteins that are continuous from their N-terminal extracellular domains through to their C-terminal cytoplasmic domains (Fig. 3, Fig. 12 and Fig. 13), the tandem repeat array containing MUC1 protein is proteolytically cleaved in its extracellular domain [Ligtenberg, et al., "Cell Associated Episialin Is a Complex Containing Two Proteins Derived From a Common Precursor," *J. Biol. Chem.*, Vol. 267, pp. 6171-6177 (1992)]. Integrity of the MUC1 extracellular domain as in the MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt proteins is likely to be essential for ligand binding.

Furthermore, the MUC1 amino acid sequence reveals striking similarities to sequences in the extracellular domain of cytokine receptors that are known to participate in ligand binding. Significantly, this homology maps in close proximity to the region where proteolytic cleavage

- 10 -

occurs in the tandem repeat array containing MUC1 protein, suggesting that integrity of this site in the MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt proteins is of prime importance for both ligand binding and signal transmission. This demonstrates that the MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt proteins are cytokine-like receptor molecules.

Furthermore, experiments carried out with the MUC1 proteins described previously in the literature and which are characterized by the presence of the tandem repeat array, showed that these proteins do not transform cells into cancerous cells, and specifically when expression vectors containing cDNA coding for the tandem repeat array MUC1 protein were transfected into eucaryotic cells, the said transfectants did not become tumorigenic. In contradistinction thereto, transfection of expression vectors containing cDNA coding for the MUC1/Y protein of the present invention into cells, caused the said cells to become tumorigenic, as described hereinbelow.

As is known, the biological effects of many factors controlling cell proliferation, differentiation and metabolism are mediated by membrane-located proteins (receptors) that participate in signal transduction processes. Invariably, growth factor binding to specific cell surface receptors initiates a signalling cascade that is transduced in many cases via phosphorylation of tyrosine residues within the receptor protein [M.J. Pazijn and L.T. Williams, "Triggering Signalling Cascades by Receptor Tyrosine Kinases," TIBS, Vol. 17, pp. 374-378 (1992)]. Assembly of receptor signalling complexes formed between the receptor protein and SRC homology 2 (SH2) domain containing proteins that interact with phosphorylated tyrosine residues

- 11 -

present in the receptor cytoplasmic domain, mediates the signal transduction process. This triggering ultimately results in the activation of specific gene expression involving transcription of both immediate and delayed response genes.

A number of cell surface receptor proteins are likely involved in both the origin and progression of human breast cancer - a prime example is the neu (erbB-2) membrane located receptor molecule [D.J. Slamon, et al., "Studies on the HER-2/neu Protooncogene in Human Breast and Ovarian Cancer," Science, Vol. 244, pp. 707-712 (1989)]. It is therefore unfortunate to note, however, that only exceptionally few genes that code for signal transducing molecules in general, and membrane-located receptor proteins in particular, have to date been implicated in the development of human breast cancer.

Thus, as stated above, there have now been identified and characterized novel protein products of the MUC1 gene, designated herein as MUC1/X, MUC1/Y and MUC1/V, that reside in the cell membrane and function as receptor proteins, and are highly expressed in human breast cancer tissue. There have also now been identified and characterized novel protein products of the MUC1 gene, designated herein as MUC1/W and MUC1/Z, the latter of which has been found to function as ligands, and the former of which is believed to have a similar function, based on its structure.

These proteins and the /alt configurations thereof, as well as functional derivatives thereof, form the basis of the present invention.

- 12 -

Thus, the present invention further provides a pharmaceutical composition comprising as an active ingredient therein a biochemically purified MUC1 protein selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, MUC1/W, MUC1/W/alt, MUC1/Z, MUC1/Z/alt and functional derivatives thereof, devoid of a tandem repeat array.

More specifically, the present invention provides, inter alia, a pharmaceutical composition for the treatment of human breast cancer, comprising as an active ingredient therein a biochemically pure MUC1 protein selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, MUC1/W, MUC1/W/alt, MUC1/Z, MUC1/Z/alt and functional derivatives thereof, in soluble form and in combination with a pharmaceutically acceptable carrier.

The invention also provides a conjugated toxin for the treatment of human breast cancer, comprising a MUC1 protein selected from the group consisting of MUC1/W, MUC1/W/alt, MUC1/Z, MUC1/Z/alt, and functional derivatives thereof, attached to a cytotoxic agent.

In another aspect of the present invention, there is provided a diagnostic agent for the detection of human breast cancer cells, comprising a detectable labelled MUC1 protein selected from the group consisting of MUC1/W, MUC1/W/alt, MUC1/Z, MUC1/Z/alt, and functional derivatives thereof.

The invention also provides a diagnostic agent for identification of sites in the body to which breast cancer cells have spread, comprising a detectable labelled MUC1 protein selected from the group consisting of MUC1/W,

- 13 -

MUC1/W/alt, MUC1/Z, MUC1/Z/alt, and functional derivatives thereof.

As will be realized from the above, the invention also includes a method for the treatment of human breast cancer, comprising administering to an individual having human breast cancer cells an amount of soluble MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, or MUC1/V/alt receptors, sufficient to inhibit the binding of MUC1 ligands to said cells.

In yet another aspect of the present invention, there is provided a method for the treatment of human breast cancer, comprising administering to an individual having human breast cancer cells an amount of a ligand-toxin conjugant comprising a ligand selected from MUC1/W, MUC1/W/alt, MUC1/Z or MUC1/Z/alt, fused to a cytotoxic toxin.

The MUC1/Z and MUC1/W proteins may be used:

- a) for breast cancer diagnosis and prognosis, both in vivo and in vitro;
- b) for imaging cancer tissue; and
- c) for therapy of breast cancer patients.

Breast Cancer Diagnosis and Prognosis

As the MUC1/W and MUC1/Z proteins are synthesized by breast cancer tissue and are secreted from the cell, their serum levels can serve as markers for the disease. Assays employing antibodies directed against the MUC1/W and MUC1/Z proteins are used to analyze the serum levels of these proteins. This provides a means for diagnosing individuals with early breast cancer, and/or for monitoring the

- 14 -

progression of breast cancer in patients who already have been diagnosed.

In general, ELISAs are the preferred immunoassays employed to assess the amount of the new proteins described and claimed herein present in a specimen. ELISA assays are well-known to those skilled in the art. Both polyclonal and monoclonal antibodies can be used in the assays. Where appropriate, other immunoassays, such as radioimmunoassays (RIA) can be used, as known to those skilled in the art. Available immunoassays are extensively described in the patent and scientific literature. See, for example, U.S. Patents 3,791,932; 3,839,153; 3,850,752; 3,850,578; 3,853,987; 3,867,517; 3,879,262; 3,901,654; 3,935,074; 3,984,533; 3,996,345; 4,034,074 and 4,098,876, as well as Sambrook, et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York, U.S.A. (1989), and E. Harlow and D. Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York, U.S.A. (1988).

Imaging of Breast Cancer Tissue

The identification of sites in the body to which breast cancer cells have spread is of prime importance for the successful eradication of the disease. The MUC1/Z ligand specifically homes in onto breast cancer cells expressing the target MUC1/X, MUC1/Y and MUC1/V receptor molecules, providing the means for efficiently localizing cancerous tissue. Imaging is performed by tagging the MUC1/Z ligand with, for example, radioactivity, injecting the labelled MUC1/Z protein into the patient, and monitoring its localization within the body.

- 15 -

Therapy of Breast Cancer Patients with Ligand

1. Ligand as a Drug-Delivery System

Using the MUC1/Z ligand as a drug delivery system, ligand-toxin conjugates are prepared, such as MUC1/Z fused to a cytotoxic toxin.

The toxin thus specifically homes in onto the target breast cancer cell, which is then killed. Alternatively, the ligand is labelled with cytotoxic levels of radioactivity. The target breast cancer cells are then directly eradicated by the radioactively- labelled ligand.

2. Blockade of MUC1/X, MUC1/Y and MUC1/V Receptors without Receptor Activation

By using defined regions of the ligand that only bind to the receptor, yet do not activate it, it is possible to effectively "swamp" the receptors present on the breast cancer cell with non-activating ligand. Receptor occupancy with non-activating ligands (antagonistic ligands) will preclude the binding of activating ligands, thereby limiting the growth of the breast cancer cell.

The specification and claims provide guidance for the use of the invention in humans. The Investigator's Handbook provided by the Cancer Therapy Evaluation Program, Division of Cancer Treatment, National Cancer Institute, U.S.A., indicates that the starting dose for Phase I trials is based on animal data such as rodent equivalent LD₅₀. Further, the manual (page 22) indicates that animal studies carried out prior to Phase I trials provide the investigator with a prediction of the likely effects. [See also J.S. Driscoll, "The Preclinical New Drug Research Program of the National Cancer Institute," Cancer Treatment Reports, Vol. 68,

- 16 -

pp. 63-76 (1984).] Therefore, the data accumulated in a mouse model is not only acceptable in determining human doses and protocols, but is considered highly predictive.

The new MUC1 proteins of the present invention, i.e., the proteins selected from the group of proteins consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, MUC1/W, MUC1/W/alt, MUC1/Z and MUC1/Z/alt, as well as their functional derivatives as defined herein, are prepared by recombinant DNA technology and polypeptide synthesis.

Thus, the new MUC1 proteins of the present invention are prepared by culturing a host cell transformed with an expression vector comprising DNA encoding an amino acid sequence of the new MUC1 proteins in a nutrient medium, and recovering the new MUC1 proteins from the cultured broth.

Particulars of the above-mentioned process are explained in detail below.

The host cell may include a microorganism [bacteria (e.g., Escherichia coli, Bacillus subtilis, etc.); yeast (e.g., Saccharomyces cerevisiae, etc.)], cultured human or animal cells (e.g., CHO cell, L929 cell, etc.), cultured plant cells, and cultured insect cells. Preferred examples of the microorganism include bacteria, especially a strain belonging to the genus Escherichia (e.g., E. coli HB-101, ATCC 33694; E. coli HB-101-16, FERM BP-1872; E. coli 294, ATCC 31446; E. coli X-1776, ATCC 31537, etc.); yeast, animal cell lines (e.g., mouse L929 cell, Chinese hamster ovary (CHO) cell, etc.), and the like.

- 17 -

When the bacterium, especially E. coli, is used as a host cell, the expression vector usually comprises at least a promoter-operator region, initiation codon, DNA encoding the amino acid sequence of the new MUC1 proteins, termination codon, terminator region, and replicatable unit. When yeast or an animal cell is used as host cell, the expression vector is preferably composed of at least promoter, initiation codon, DNA encoding the amino acid sequence of the signal peptide and the new MUC1 proteins, and termination codon, and it is possible that enhancer sequences, 5'- and 3'-noncoding region of the native MUC1 proteins, splicing junctions, polyadenylation site and replicatable unit are also inserted into the expression vector.

The promoter-operator region comprises promoter, operator and Shine-Dalgarno (SD) sequence (e.g., AAGG, etc.). Examples of the promoter-operator region include conventionally employed promoter-operator region (e.g., lactose-operon, PL-promoter, trp-promoter, etc.) and the promoter for the expression of the new MUC1 protein in mammalian cells may include HTLV-promoter, SV40 early- or late-promoter, LTR-promoter, mouse metallothionein I (MTT)-promoter and vaccinia-promoter.

Preferred initiation codon includes methionine codon (ATG).

The DNA encoding signal peptide includes the DNA encoding signal peptide of the new MUC1 proteins.

The DNA encoding the amino acid sequence of the signal peptide or the new MUC1 proteins is prepared in a conventional manner, such as a partial or whole DNA

- 18 -

synthesis using DNA synthesizer and/or treatment of the complete DNA sequence coding for native or mutant MUC1 proteins inserted in a suitable vector obtainable from a transformant or genome in a conventional manner (e.g., digestion with restriction enzyme, dephosphorylation with bacterial alkaline phosphatase, ligation using T4 DNA ligase).

The termination codon(s) include conventionally employed termination codon (e.g., TAG, TGA, etc.).

The terminator region contains natural or synthetic terminator (e.g., synthetic fd phage terminator, etc.).

The replicatable unit is a DNA sequence capable of replicating the whole DNA sequence belonging thereto in the host cells and includes natural plasmid, artificially modified plasmid (e.g., DNA fragment prepared from natural plasmid) and synthetic plasmid, and preferred examples of the plasmid include plasmid pBR 322 or artificially modified plasmid thereof (DNA fragment obtained from a suitable restriction enzyme treatment of pBR 322) for E. coli; plasmid pRSVneo ATCC 37198, plasmid pSV2dhfr ATCC 37145, plasmid pdBPV-MMTneo ATCC 37224, plasmid pSV2neo ATCC 37149 for mammalian cell.

The enhancer sequence includes the enhancer sequence (72 bp) of SV40.

The polyadenylation site includes the polyadenylation site of SV40.

The splicing junction includes the splicing junction of SV40.

- 19 -

The promoter-operator region, initiation codon, DNA encoding the amino acid sequence of the new MUC1 proteins, termination codon(s) and terminator region are consecutively and circularly linked together with an adequate replicatable unit (plasmid) if desired, using adequate DNA fragment(s) (e.g., linker, other restriction site, etc.) in a conventional manner (e.g., digestion with restriction enzyme, phosphorylation using T4 polynucleotide kinase, ligation using T4 DNA ligase) to give an expression vector. When mammalian cell line is used as a host cell, it is possible that enhancer sequence, promoter, 5'-noncoding region of the cDNA of the native MUC1 proteins, initiation codon, DNA encoding amino acid sequences of the signal peptide and the new MUC1 termination codon(s), 3'-noncoding region, splicing junctions and polyadenylation site are consecutively and circularly linked together with an adequate replicatable unit in the above manner.

The expression vector is inserted into a host cell by methods known per se. The insertion is carried out in a conventional manner (e.g., transformation including transfection, microinjection, etc.) to give a transformant including transfectant.

For the production of the new MUC1 proteins in the process of the present invention, thus obtained transformant comprising the expression vector is cultured in a nutrient medium.

The nutrient medium contains carbon source(s) (e.g., glucose, glycerine, mannitol, fructose, lactose, etc.) and inorganic or organic nitrogen source(s) (e.g., ammonium sulfate, ammonium chloride, hydrolysate of casein, yeast extract, polypeptone, bactotrypton, beef extracts, etc.). If

- 20 -

desired, other nutritious sources [e.g., inorganic salts (e.g., sodium or potassium biphosphate, dipotassium hydrogen phosphate, magnesium chloride, magnesium sulfate, calcium chloride), vitamins (e.g., vitamin B1), antibiotics (e.g., ampicillin), etc.] are added to the medium. For the culture of mammalian cell, Dulbecco's Modified Eagle's Minimum Essential Medium (DMEM) supplemented with fetal calf serum and an antibiotic is often used.

The culture of transformant is generally be carried out at pH 5.5-8.5 (preferably pH 7-7.5) and 18-40°C (preferably 25-38°C) for 5-50 hours.

When a bacterium such as E. coli is used as a host cell, thus produced new MUC1 proteins generally exist in cells of the cultured transformant and the cells are collected by filtration or centrifugation, and cell wall and/or cell membrane thereof are destroyed in a conventional manner (e.g., treatment with supersonic waves and/or lysozyme, etc.) to give debris. From the debris, the new MUC1 proteins are purified and isolated in a conventional manner, as generally employed for the purification and isolation of natural or synthetic proteins [e.g., dissolution of protein with an appropriate solvent (e.g., 8M aqueous urea, 6M aqueous guanidium salts, etc.), dialysis, gel filtration, column chromatography, high performance liquid chromatography, etc.]. When a mammalian cell is used as a host cell, the produced new MUC1 proteins generally exist in the culture solution. The culture filtrate (supernatant) is obtained by filtration or centrifugation of the cultured broth. From the culture filtrate, the new MUC1 proteins are purified in a conventional manner.

- 21 -

As will be realized, having now identified the new MUC1 proteins of the present invention, purified antibodies, both polyclonal and monoclonal, which specifically bind respectively to each of said proteins can be readily prepared by methods per se known in the art. Once said antibodies are prepared, they can be conjugated to a therapeutic drug or a detectable moiety and/or bound to a solid support.

The preparation of said antibodies also enables the carrying-out of a bioassay for determining the amount of a MUC1 protein selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, MUC1/W, MUC1/W/alt, MUC1/Z and MUC1/Z/alt or a functional derivative thereof devoid of a tandem repeat array, comprising (a) contacting the biological sample with an antibody under conditions such that a specific complex of the antibody and said MUC1 protein can be formed; and (b) determining the amount of the antibody/MUC1 protein complex, the amount of the complex indicating the amount of said MUC1 protein in the biological sample, and allows the method of detecting the presence of a cancer in a subject comprising determining the presence of a detectable amount of said MUC1 protein in a biopsy from the subject, the presence of a detectable amount of said MUC1 protein relative to the absence of MUC1 protein in a normal control indicating the presence of a cancer, and the method of determining the prognosis of a subject having cancer, comprising determining the presence of a detectable amount of said MUC1 protein in a biopsy from the subject, the presence of a detectable amount of MUC1 protein relative to the absence of said MUC1 protein in a normal control indicating a decreased chance of long-term survival.

- 22 -

While the invention will now be described in connection with certain preferred embodiments in the following examples and with reference to the following illustrative figures so that aspects thereof may be more fully understood and appreciated, it is not intended to limit the invention to these particular embodiments. On the contrary, it is intended to cover all alternatives, modifications and equivalents as may be included within the scope of the invention as defined by the appended claims. Thus, the following examples which include preferred embodiments of the novel proteins, the functional derivatives thereof, the combination thereof with cytotoxic agents and detectably labelled markers, as well as the preparation of DNA constructs, vectors, and transfected hosts encoding and incorporating the same, and the various uses thereof, will serve to illustrate the practice of this invention, it being understood that the particulars shown are by way of example and for purposes of illustrative discussion of preferred embodiments of the present invention only and are presented in the cause of providing what is believed to be the most useful and readily understood description of formulation procedures as well as of the principles and conceptual aspects of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

In the drawings:

- Fig. 1A is a scheme of alternative splice events (W, X, Y and Z) that delete the MUC1 tandem repeat array and flanking sequences;
- Fig. 1B is a scheme of alternative splice events (W, X, Y, and Z) and nucleotide sequence of the regions 5' flanking the AG consensus splice acceptor site;

- Fig. 2 shows amino terminal amino acid sequences of the MUC1 proteins, demonstrating the two variant MUC1 signal peptide forms and sites of signal peptide cleavage;
- Fig. 3 is a scheme of the repeat array containing MUC1 protein (upper molecule) and the novel MUC1/W, MUC1/X, MUC1/Y and MUC1/Z proteins generated by alternative splicing;
- Fig. 4 is a scheme of the repeat array containing MUC1/alt protein that has the variant signal peptide at its N-terminal and the novel MUC1/Y/alt, MUC1/X/alt, MUC1/W/alt and MUC1/Z/alt proteins generated by alternative splicing;
- Fig. 5A shows the amino acid sequence of the MUC1/X protein;
- Fig. 5B shows the amino acid sequence of the MUC1/X/alt protein;
- Fig. 6A shows the amino acid sequence of the MUC1/Y protein;
- Fig. 6B shows the amino acid sequence of the MUC1/Y/alt protein;
- Fig. 6C shows the amino acid sequence of the MUC1/V protein;
- Fig. 6D shows the amino acid sequence of the MUC1/V/alt protein;
- Fig. 7A shows the amino acid sequence of the MUC1/W protein;
- Fig. 7B shows the amino acid sequence of the MUC1/W/alt protein;
- Fig. 8A shows the amino acid sequence of the MUC1/Z protein;
- Fig. 8B shows the amino acid sequence of the MUC1/Z/alt protein;
- Fig. 9 illustrates the overexpression of the novel MUC1/X, MUC1/Y, and MUC1/V proteins in human breast cancer tissue and post-translational modification by phosphorylation;
- Fig. 10 illustrates phosphorylation on tyrosine residues of the MUC1/Y protein;

- 24 -

- Fig. 11 depicts the binding of tyrosine phosphorylated MUC1 cytoplasmic domain to SH2 domains;
- Fig. 12 is a scheme depicting the repeat array containing MUC1 protein (upper drawing) and the novel MUC1/Y protein (lower drawing);
- Fig. 13 is a scheme depicting the location of tyrosine and cysteine residues in the MUC1 proteins; and
- Fig. 14 is a comparison scheme of MUC1 sequences and sequences known to interact with SH2 domains.

DETAILED DESCRIPTION OF THE INVENTION

With regard to the attached drawings, the following is a more detailed description thereof, so that the same can be more readily understood:

Fig. 1A: Scheme of alternative splice events (W, X, Y and Z) that delete the MUC1 tandem repeat array and flanking sequences. The MUC1 genomic sequence is indicated by the continuous line. The various splice events (W, X, Y and Z) that delete the tandem repeat array are indicated. The dinucleotides at the splice donor and splice acceptor sites are indicated by GT and AG, respectively. The X and Y splices retain the same reading frame (RF) as the MUC1 protein, whereas W and Z change the reading frame. The signal peptide and the transmembrane domains are indicated by SIG and TM, respectively.

Fig. 1B: Scheme of alternative splice events (W, X, Y and Z) and 5' sequences flanking the splice acceptor site. The pyrimidine-rich sequences 5' flanking the W, X, Y and Z splice acceptor sites are shown. Other symbols are as in Fig. 1A.

- 25 -

Fig. 2: Alternative MUC1 N-terminal signal peptide sequences. The amino terminal (N-terminal) amino acid sequence is presented using the one letter code. The lower sequence represents the N-terminal sequence that includes an extra 9 amino acids (boxed sequence) that is generated by an alternative splice event. Numbers appearing above the amino acid sequence represent the probability (calculated according to the Von Heijne signal peptide cleavage rules; arbitrary units are used) of signal peptide cleavage occurring at that site. The upward-facing arrow represents the most likely site of signal peptide cleavage.

Fig. 3: Scheme of the repeat array containing MUC1 protein (upper molecule) and the novel MUC1/Y, MUC1/X, MUC1/W and MUC1/Z proteins. The novel MUC1/Y, MUC1/X, MUC1/W and MUC1/Z proteins are generated by alternative splicing events that delete the central tandem repeat array (compare upper and lower molecules). All MUC1 forms contain a hydrophobic N-terminal signal sequence (slashed box at left of figure) that is co-translationally cleaved (arrow at left of figure). This is followed by the tandem repeat array (upper molecule) that is illustrated by the block of closely-spaced vertical lines. The highly hydrophobic 28 amino acid stretch constituting the transmembrane domain (TM) is shown at the C-terminal end of both MUC1 proteins, followed by the cytoplasmic domain (CYT). The region comprising the proteolytic cleavage site [Ligtenberg, et al., *J. Biol. Chem.*, *ibid.* (1992)] of the repeat array containing MUC1 protein (upper molecule) is indicated by the two vertical dotted lines just N-terminal to the transmembrane domain. Potential N-linked glycosylation sites are shown with an asterisk (*). The W and Z splice events alter the reading frame of the MUC1 protein downstream to

- 26 -

their respective splice acceptor sites, and therefore contain downstream amino acid sequences that differ from the MUC1/Y and MUC1/X proteins.

Fig. 4: Scheme of the repeat array containing MUC1/alt protein that has the variant signal peptide at its N-terminal and the novel MUC1/Y/alt, MUC1/X/alt, MUC1/W/alt and MUC1/Z/alt proteins generated by alternative splicing. The altered N-terminal (see Fig. 2) resulting from the altered signal peptide is illustrated immediately distal to the slashed box at the N-terminus. All the resulting novel MUC1/Y/alt, MUC1/X/alt, MUC1/W/alt and MUC1/Z/alt proteins will accordingly have the variant N-terminus. Other symbols are as in Fig. 3.

Fig. 5A: Amino acid sequence of the MUC1/X protein. The amino acid sequence (one letter code) is shown beneath the nucleotide sequence and begins with the initiating methionine residue. Numbering of the amino acid residues is shown to the right of the figure.

Fig. 5B: Amino acid sequence of the MUC1/X/alt protein. The amino acid sequence (one letter code) is shown beneath the nucleotide sequence and begins with the initiating methionine residue. Numbering of the amino acid residues is shown to the right of the figure.

Fig. 6A: Amino acid sequence of the MUC1/Y protein. The amino acid sequence (one letter code) is shown beneath the nucleotide sequence and begins with the initiating methionine residue. Numbering of the amino acid residues is shown to the right of the figure.

Fig. 6B: Amino acid sequence of the MUC1/Y/alt protein. The amino acid sequence (one letter code) is shown beneath the nucleotide sequence and begins with the initiating methionine residue. Numbering of the amino acid residues is shown to the right of the figure.

Fig. 6C: Amino acid sequence of the MUC1/V protein. The amino acid sequence (one letter code) is shown beneath the nucleotide sequence and begins with the initiating methionine residue. Numbering of the amino acid residues is shown to the right of the figure.

Fig. 6D: Amino acid sequence of the MUC1/V/alt protein. The amino acid sequence (one letter code) is shown beneath the nucleotide sequence and begins with the initiating methionine residue. Numbering of the amino acid residues is shown to the right of the figure.

Fig. 7A: Amino acid sequence of the MUC1/W protein. The amino acid sequence (one letter code) is shown beneath the nucleotide sequence and begins with the initiating methionine residue. Numbering of the amino acid residues is shown to the right of the figure.

Fig. 7B: Amino acid sequence of the MUC1/W/alt protein. The amino acid sequence (one letter code) is shown beneath the nucleotide sequence and begins with the initiating methionine residue. Numbering of the amino acid residues is shown to the right of the figure.

Fig. 8A: Amino acid sequence of the MUC1/Z protein. The amino acid sequence (one letter code) is shown beneath the nucleotide sequence and begins with the initiating

- 28 -

methionine residue. Numbering of the amino acid residues is shown to the right of the figure.

Fig. 8B: Amino acid sequence of the MUC1/Z/alt protein. The amino acid sequence (one letter code) is shown beneath the nucleotide sequence and begins with the initiating methionine residue. Numbering of the amino acid residues is shown to the right of the figure.

Fig. 9: Overexpression of the novel MUC1/X, MUC1/Y and MUC1/V proteins in human breast cancer tissue and post-translational modification by phosphorylation.

(A) Cell lysates prepared from breast cancer cells (lane 2), primary human breast cancer tissues from 3 different patients (lanes 1, 4 and 5) and the adjacent normal breast tissues (lanes 3 and 6), were analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE), transferred to nitrocellulose and immunoblotted with a rabbit polyclonal antibody directed against the MUC1 cytoplasmic domain. The regions of specific immunoreactivity are indicated by the 3 open arrows to the left of the figure.

(B) The novel MUC1/Y protein may be post-translationally modified by phosphorylation. Radioactive inorganic phosphate (^{32}P) was added to stable Ras transformed 3T3 cell transfectants expressing the MUC1/Y protein and following a 5-hour incubation the cells were lysed. Cell lysates subjected to immunoprecipitation with either pre-immune serum or with immune serum generated against the 62 C-terminal amino acids of the MUC1 cytoplasmic domain (lanes 1 and 2, respectively) were analyzed by SDS-PAGE, followed by autoradiography. The phosphorylated MUC1/Y protein is clearly visible in lane 2

(arrow to the right of the figure). Molecular size standards are indicated at left of figures in kilodaltons.

Fig. 10: Phosphorylation on tyrosine residues of the MUC1/Y protein. The immunoprecipitated phosphorylated MUC1 proteins [from lane 2 in Fig. 9(B)] were isolated from SDS-acrylamide (10%) gel and hydrolyzed in 6M HCl at 110°C for 1 hour. Labelled phosphoaminoacids (with added unlabelled internal phosphoamino acid markers) were analyzed by thin-layer high voltage electrophoresis, followed by Phosphorimager analysis. The position of migration of phosphoserine, phosphothreonine and phosphotyrosine are indicated by PS, PT and PY respectively, and inorganic phosphate is shown by Pi.

Fig. 11: Binding of tyrosine phosphorylated MUC1 cytoplasmic domain to SH2 domains.

(A) The complete 72 amino acid sequence of the human MUC1 cytoplasmic domain is shown, using the one letter amino acid code. Indicated below this are changes in the mouse MUC1 homologue. The 7 tyrosine residues in the cytoplasmic domain are highlighted with an asterisk, and likely sites of interaction between phosphotyrosine-containing peptide sequences (boxed regions within the cytoplasmic domain amino acid sequence) and SH2 domain containing proteins (boxed at the bottom of the figure) are shown. The cysteine-containing sequence is circled at the N-terminal of the cytoplasmic domain.

(B) Recombinant MUC1 cytoplasmic domain was synthesized as a fusion protein with N-terminal DHFR protein (from *Halobacterium*) using the pET system. The gel purified recombinant protein was in-vitro tyrosine phosphorylated by incubation with gamma 32 P-ATP and highly purified EGF receptor (EGF-R) protein isolated from A431 cells. The

- 30 -

radioactively-labelled MUC1 cytoplasmic domain was repurified from a SDS-acrylamide (10%) gel and incubated overnight at 4°C, with either GST (glutathione transferase) beads alone (lane 1), or with GST/GRB-2 fusion protein beads (GRB-2, lane 2). The beads were then extensively washed and labelled bound proteins analyzed by SDS-PAGE. Specific GRB-2 binding of labelled MUC1 cytoplasmic domain is indicated by the arrow to the right of the figure.

(C) Labelled tyrosine phosphorylated MUC1 cytoplasmic domain, purified by SDS acrylamide (10%) gel, was incubated with agarose beads bound to src SH2 domain (src, lane 1), the C-terminal p85 phosphatidyl inositol (PI) 3' kinase SH2 domain (PI, lane 2), and the N-terminal phospholipase C gamma 1 SH2 domain (lip. C, lane 3) and analyzed as described above. Specific binding to the src and phospholipase C SH2 domains (lanes 1 and 3, respectively) is indicated by the arrow to the right of the figure. No binding was observed to the C-terminal p85 (PI) 3' kinase SH2 domain (lane 2).

Fig. 12: Scheme showing the repeat array containing MUC1 protein (upper drawing) and the novel MUC1/Y protein (lower drawing). The novel MUC1/Y form is generated by an alternative splicing event that deletes the central tandem repeat array (compare upper and lower molecules). Both MUC1 forms contain a hydrophobic N-terminal signal sequence (slashed box at left of figure) that is co-translationally cleaved (arrow at left of figure). This is followed by the tandem repeat array (upper molecule) that is illustrated by the block of closely-spaced vertical lines. The highly hydrophobic 28 amino acid stretch constituting the transmembrane domain (TM) is shown at the C-terminal end of both MUC1 proteins, followed by the cytoplasmic domain (CYT). The region comprising the proteolytic cleavage site

- 31 -

[Ligtenberg, et al., J. Biol. Chem., *ibid.* (1992)] of the repeat array containing MUC1 protein (upper molecules) is indicated by the two vertical dotted lines just N-terminal to the transmembrane domain. The regions recognized by the anti-repeat and anti- cytoplasmic domain (anti-cyt) antibodies are indicated and potential N-linked glycosylation sites are shown with an asterisk (*).

Fig. 13: Scheme showing the location of tyrosine and cysteine residues in the MUC1 proteins. The location of tyrosine and cysteine residues are indicated above the rectangles by vertical lines and asterisks, respectively. Both MUC1 forms contain a hydrophobic N-terminal signal sequence (slashed box at left of figure) that is co-translationally cleaved (arrow at left of figure). This is followed by the tandem repeat array (upper molecule) that is illustrated by the block of closely-spaced vertical lines. The highly hydrophobic 28 amino acid stretch constituting the transmembrane domain (TM) is shown at the C-terminal end of both MUC1 proteins, followed by the cytoplasmic domain (CYT). The region comprising the proteolytic cleavage site [Ligtenberg, et al., J. Biol. Chem., *ibid.* (1992)] of the repeat array containing MUC1 protein (upper molecule) is indicated by the two vertical dotted arrows just N-terminal to the transmembrane domain. The regions recognized by the anti-cytoplasmic domain (anti-cyt) antibodies are indicated.

Fig. 14: Phosphotyrosine-Containing Peptide Sequences Recognized by SH2 Domains and Their Comparison with MUC1 Cytoplasmic Domain Sequences. The sequence specificity of the peptide-binding sites of SH2 domains has been previously determined using a phosphopeptide library [Songyang, et al., Cell, Vol. 72, pp. 767-778 (1993)] and the data presented in

- 32 -

this Figure are in part from Table 3 of that reference. The preferred amino acids 1, 2 and 3 residues C-terminal to phosphotyrosine are indicated in the columns labelled pY + 1, pY + 2 and pY + 3. The top line in each group relates to the most preferred sequence, with lowered preferences in the second and third lines. The boxed sequences correlate best with MUC1 cytoplasmic domain sequences that are indicated in the right-hand column.

Experimental work detailed below has unequivocally demonstrated that:

- a) the MUC1/X, MUC1/Y and MUC1/V proteins are highly and differentially expressed in breast cancer tissue as compared to normal breast tissue [see Fig. 9];
- b) the MUC1/X, MUC1/Y and MUC1/V proteins are extensively phosphorylated [see Fig. 9];
- c) phosphorylation occurs almost exclusively on tyrosine residues [see Fig. 10];
- d) the phosphorylated MUC1/X, MUC1/Y and MUC1/V proteins interact specifically with the SRC-homology (SH) domain SH2- and SH3-containing proteins, GRB-2, SRC and phospholipase C gamma-1 [Fig. 11]; and
- e) the MUC1/X, MUC1/Y and MUC1/V proteins potentiate the transformed phenotype of cells and significantly enhance the in-vivo tumorigenic potential of mammary epithelial cells.

This experimental data demonstrates that the MUC1/X, MUC1/Y and MUC1/V proteins function as cell surface receptor molecules participating in signal transduction, and are

- 33 -

intimately related to the development of human breast cancer.

To assess expression of the MUC1 proteins in-vivo, extracts of human tissue samples were run on SDS denaturing gels, transferred and probed with polyclonal antibodies directed against the MUC1 cytoplasmic domain. Analyses were performed on malignant breast tumor tissue samples [Fig. 9A, lanes 4 and 5], together with extracts from breast tissue adjacent to the biopsied tumor sample [Fig. 9A, lanes 3 and 6]. Little or no specific immunoreactivity was observed in the non-malignant breast tissue samples [Fig. 9A, lanes 3 and 6].

In marked contrast thereto, proteins specifically reactive with the anticytoplasmic domain antibodies were highly expressed both in breast cancer cells grown in-vitro and in the primary breast cancer tissue samples [Fig. 9A, lanes 2, 4 and 5 respectively].

The immunoreactive proteins migrated to distinct positions correlating to molecular masses of approximately 25-30, 35 [in the in-vitro grown breast cancer cells, lane 2], and 40-43 kDa. Some of these immunoreactive proteins may be generated by proteolytic cleavages occurring on the large polymorphic tandem repeat array containing MUC1 protein at positions N-terminal to the transmembrane domain [Fig. 12, upper molecule, the two dotted arrows just N-terminal to the transmembrane domain]. However, the MUC1/X, MUC1/Y proteins [Fig. 12, lower molecule], and MUC1/V proteins are also likely represented by one or more of these immunoreactive proteins. In distinguishing between these possibilities, we were considerably aided by the identification of a third breast tumor tissue sample

- 34 -

[Fig. 9, lane 1], that expresses specific anticytoplasmic domain immunoreactive proteins with molecular masses of approximately 40-43 kDa and 35 kDa [compare Fig. 9, lanes 1 and 2]. Probing an identical immunoblot with monoclonal antibodies that recognize an epitope contained within the tandem repeat array, showed high levels of expression of the large polymorphic MUC1 proteins in the breast cancer cell samples correlating to lanes 2, 4 and 5 - no immunoreactive proteins corresponding to the large polymorphic MUC1 proteins were detected in the third breast tumor correlating to lane 1 [data not shown]. These data suggest therefore that this third breast tumor tissue solely expresses the MUC1/X, MUC1/Y and MUC1/V protein forms and thereby indicate that the 35 and 40-43 kDa immunoreactive proteins are in fact the MUC1/X and MUC1/Y proteins.

Tyrosine Phosphorylation of the MUC1/X, MUC1/Y and MUC1/V Proteins

The calculated molecular mass of the MUC1/Y protein, as determined by its primary amino acid sequence, is 25,986 Daltons. An increase in the molecular mass of the MUC1/Y protein [to 35 and 40-43 kDa proteins] may occur by post-translational modifications such as glycosylation and/or phosphorylation. To investigate whether the MUC1/Y protein is phosphorylated, radioactively-labelled inorganic phosphate was added to stable transfectants expressing the MUC1/Y protein, and cell lysates were subjected to anti-MUC1 cytoplasmic domain immunoprecipitation.

Specifically immunoprecipitated MUC1/Y protein migrated with a molecular mass of 40-43 kDa, and demonstrated a prominent signal [Fig. 9B, lane 2], indicating that the 40-43 kDa MUC1/Y proteins [Fig. 9] are phosphorylated

- 35 -

proteins. A phosphoamino acid analysis performed on the isolated phosphorylated MUC1/Y protein shows that greater than 90% of the phosphorylation occurs on tyrosine residues, with much reduced levels of phosphoserine and almost undetectable levels of threonine phosphorylation [Fig. 10].

Considering that within the cell greater than 99% of total protein phosphorylation occurs solely on serine and threonine residues, the almost exclusive tyrosine phosphorylation of the MUC1/Y protein is especially striking. Phosphorylated tyrosine residues play a pivotal role in signal transduction pathways [M.J. Pazin and L.T. Williams, *ibid.* (1992)] as, for example, those initiated by growth factor receptors such as epidermal growth factor receptor (EGF-R), platelet derived growth factor receptor (PDGF-R), colony stimulating factor-1 receptor (CSF1-R), etc. This suggests therefore, that the extensively tyrosine phosphorylated MUC1/Y protein may also be performing an important signal-transducing function.

MUC1/Y Protein Interaction With SH2 Domain Proteins

Analysis of the MUC1 proteins demonstrates the following features:

- 1) biased localization of tyrosine residues in the cytoplasmic domain and sequences N-terminal to it [Fig. 13];
- 2) all tyrosine residues within the polymorphic MUC1 proteins are retained in the MUC1/X, MUC1/Y and MUC1/V proteins [Fig. 13];

- 36 -

3) extensive similarity between the human and mouse MUC1 proteins within the amino acid MUC1 cytoplasmic domain [Fig. 11]; and

4) marked similarity between tyrosine-containing sequences located within the MUC1 cytoplasmic domain and phosphotyrosine-containing peptide sequences that are recognized by SH2 domain-containing proteins [Fig. 11].

Bearing in mind that the MUC1/X, MUC1/Y and MUC1/V proteins are extensively phosphorylated on tyrosine residues, these remarkable features indicate that the MUC1/X, MUC1/Y and MUC1/V proteins act as receptor-like molecules that participate in signal transduction. Thus, it is now believed that the cytoplasmic domain of the MUC1/X, MUC1/Y and MUC1/V proteins acts as a "surrogate" kinase insert, in a way similar to CD19 [D.A. Tuveson, et al., "CD19 of B Cells as a Surrogate Kinase Insert Region to Bind Phosphatidylinositol 3-Kinase," Science, Vol. 260, pp. 986-988 (1993)], and undergoes transphosphorylation on tyrosine residues by other activated tyrosine kinases with which it may specifically interact. This then forms a signalling complex composed of the phosphorylated MUC1/X, MUC1/Y and MUC1/V proteins and SH2 domain-containing proteins [C.A. Koch, et al., "SH2 and SH3 Domains: Elements that Control Interactions of Cytoplasmic Signalling Proteins," Science, Vol. 252, pp. 668-674 (1991)], thereby initiating signal transduction.

To test whether the cytoplasmic domain of the MUC1/Y protein has the potential to interact specifically with SH2 domain-containing proteins, recombinant MUC1 cytoplasmic domain was synthesized and radioactively phosphorylated on

- 37 -

its tyrosine residues with highly purified epidermal growth factor receptor (EGF-R). Incubation of the phosphorylated MUC1 cytoplasmic domain with either glutathione transferase (GST) alone, or with Growth Factor Receptor Binding Protein 2 [E.J. Lowenstein, et al., "The SH2 and SH3 Domain-Containing Protein GRB2 Links Receptor Tyrosine Kinases to Ras Signalling," Cell, Vol. 70, pp. 431-442 (1992)]/GST (GRB-2/GST) fusion protein bound to agarose beads, demonstrated marked binding to the GRB-2 protein [Fig. 11B]. Analysis of the MUC1 cytoplasmic domain amino acid sequence [Fig. 11A and Fig. 14] indicates that it may also interact with additional SH2 domain-containing proteins.

Further experimentation demonstrated that purified, recombinant MUC1 cytoplasmic domain protein that had been phosphorylated on its tyrosine residues specifically bound to the SRC SH2 domain and to the SH2 domain derived from the N-terminal part of the phospholipase C gamma 1 protein [Fig. 11C, lanes 1 and 3]. Under identical conditions, no binding was observed to the C-terminal p85 phosphatidylinositol (PI) 3' kinase SH2 domain.

To validate in the in-vivo situation, findings that demonstrate in-vitro interactions of the MUC1/Y protein with multiple SH2 domain-containing proteins and, in particular, with the GRB-2 protein, human breast cancer tissue cell lysates were prepared and incubated with either GST (glutathione transferase) beads alone, or with GST/GRB-2 fusion protein beads. Bound proteins were analyzed by SDS gel electrophoresis, transferred and subjected to probing with anti-MUC1 cytoplasmic domain antibodies. The MUC1/Y protein was detected only in the sample that had been incubated with the GST/GRB-2 fusion protein beads,

indicating that in the in-vivo situation the MUC1/Y protein potentially interacts with GRB-2 protein.

MUC1/X, MUC1/Y and MUC1/V Protein Expression Alters Cell Morphology and Increases Tumorigenic Potential

As the GRB-2 protein plays a key role in connecting tyrosine kinase receptors with the ras signal transduction system [E.J. Lowenstein, et al., *ibid.* (1992)], and as shown above, the MUC1/Y proteins contact the GRB-2 protein, the effect of MUC1/Y protein expression on the morphology of ras transformed 3T3 fibroblasts was investigated. Transfectants were generated from ras transformed 3T3 fibroblasts with the neomycin resistance gene alone, and in combination with an expression vector harboring cDNA coding, for either the MUC1/Y proteins or the large tandem repeat array containing MUC1 protein. The parental ras transformed 3T3 fibroblasts, and control cells transfected only with the neomycin resistance gene, grew mostly in foci and cell clusters. As previously reported, transfectants expressing the large tandem repeat array containing MUC1 protein displayed decreased cellular aggregation and did not grow in foci; this is likely due to the known anti-adhesive properties of the tandem repeat array containing MUC1 protein. The effect of MUC1/Y protein expression on cell morphology was, however, immediately apparent. These transfectants displayed a marked increase in the number of foci, an altered phenotype that was observed in all independent MUC1/Y protein-expressing transfectants analyzed. This is indicative of the fact that expression of the MUC1/Y protein is indeed potentiating the transforming potential of the cell.

- 39 -

Next, tests were conducted to determine whether MUC1/Y protein expression alters the tumorigenic potential of mammary epithelial cells. Transfectants were generated using the DA3 mouse mammary epithelial cell line, derived from a DMBA-induced mouse mammary carcinoma, and expression of the MUC1/Y protein in the transfectants was assessed by Western blotting. Positive MUC1/Y transfectants, as well as tandem repeat array containing MUC1 transfectants and control neomycin transfectants, were injected intramuscularly into female Balb/c mice at three different cell concentrations ($5 \cdot 10^4$, 10^5 and $5 \cdot 10^6$) and the mice were monitored for tumor development.

Mice injected with transfectants expressing the tandem repeat array containing MUC1 protein, or with the control neomycin transfectants, showed similar patterns of tumor development. In marked contrast however, tumors developed rapidly in the MUC1/Y transfectant group and preceded the appearance of tumors in the other two groups by weeks to months, at all cell concentrations tested. For example, tumors developed in all mice (5 per group) injected with the MUC1/Y transfectant ($5 \cdot 10^5$ cells per mouse) only 7 days following injection. Animals injected with the control neomycin transfectants showed tumor development in three out of five mice that were first observed 6 weeks following injection. This pattern of increased tumorigenicity of the MUC1/Y transfectants was consistently observed at all other cell concentrations tested.

The experimental work described above demonstrates that the MUC1/Y proteins are highly expressed in human breast cancer tissue; are extensively phosphorylated on tyrosine residues; interact specifically with the SRC homology domain

- 40 -

(SH2) containing proteins GRB-2, SRC and phospholipase C gamma-1; and increase cellular tumorigenic potential.

As is seen from the structure of the MUC1/X molecule, it is highly similar to the MUC1/Y molecule, except for the insertion of 18 amino acids between amino acid residue numbers 53 and 54 in the MUC1/Y sequence. The MUC1/X protein is therefore believed to function as a receptor molecule in a similar fashion to the MUC1/Y protein, although its affinity for ligand may differ. This is also true for the /alt configurations of MUC1/Y and MUC1/X.

As is seen from the structure of the MUC1/V molecule, it is highly similar to the MUC1/Y molecule. The MUC1/V protein is therefore believed to function as a receptor molecule in a similar fashion to the MUC1/Y protein, although its affinity for ligand may differ. This is also true for the /alt configurations of MUC1/Y, MUC1/X and MUC1/V.

Taken together, the above data indicate that the MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V and MUC1/V/alt proteins act as signal-transducing receptor-like molecules that form a signalling complex which is intimately related to the oncogenetic process.

The MUC1/X, MUC1/Y and MUC1/V proteins are, however, different from classical receptor tyrosine kinases, in that they do not contain a catalytical tyrosine kinase domain. One of the postulates of the present hypothesis is that the cytoplasmic domains of the MUC1/X, MUC1/Y and MUC1/V proteins undergo transphosphorylation in a manner similar to that recently described for the B cell CD19 molecule [D.A.

Tuveson, et al., *ibid.* (1993)] and for other cytokine receptors.

Having identified the MUC1/X, MUC1/Y and MUC1/V receptors, it is now possible to prepare functional derivatives thereof, including purified receptors in soluble form.

Thus, e.g. by deleting sequences downstream from glycine amino acid number 173 in the MUC1/X sequence [Fig. 5A] or glycine amino acid number 155 in the MUC1/Y sequence [Fig. 6A], or glycine amino acid number 140 in the MUC1/V sequence [Fig. 6C], one produces truncated forms of the one produces truncated forms of the membrane receptors, which lack transmembrane and intracytoplasmic domains, but retain the ligand-binding extracellular portion. The affinities of soluble receptors for their ligands are comparable to those of the membrane receptors, and thus said soluble receptors can compete with the membrane bound receptors and inhibit binding of ligands to the cell and the resulting activation thereof.

Furthermore, with the molecular characterization of the MUC1/X, MUC1/Y and MUC1/V receptor molecules described herein, one can design drugs that will specifically interact with these receptors. These drugs may then be used to target breast cancer cells, either for imaging or therapeutic purposes.

Additionally, as receptor molecules are known to be shed off from cells into the peripheral circulation, assays employing antibodies directed against the MUC1/X, MUC1/Y and MUC1/V receptors can be developed to analyse the serum levels of these receptors. The serum concentrations of

these proteins, which, as previously described, are expressed at high levels in breast cancer cells, may provide a means for diagnosing individuals with early breast cancer and/or for monitoring the progression of breast cancer in patients who have already been diagnosed.

Based on the teachings of the present invention, these and other uses of the soluble receptors of the present invention will be clear to persons skilled in the art, and this especially in the light of the description and use of other soluble receptors in the literature [see, e.g., R. Fernandez-Botran, The FASEB Journal, Vol. 5, pp. 2567-2574 (1991) and S. Chamow, Int. J. Cancer, Supplement 7, pp. 69-72 (1992)].

Ligands

Receptor molecules, such as the MUC1/X, MUC1/Y and MUC1/V proteins, specifically bind ligands. The MUC1/Z protein is secreted from the cell [Figs. 3 and 4] and, as detailed below, functions as a ligand for the MUC1/X, MUC1/Y and MUC1/V receptor proteins. The MUC1/W protein is believed to have a similar ligand function, based on its structure. This is also true for the /alt configurations of MUC1/Z and MUC1/W.

By using antibodies generated in rabbits directed against MUC1/Z, we have unequivocally showed that the MUC1/Z protein is synthesized in breast tumor tissue, but not by normal breast tissue, and that it migrates in SDS-polyacrylamide gels with an apparent molecular mass of approximately 25 kDa. Binding of the 25 kDa protein to anti-MUC1/Z antibodies could be specifically competed out by the addition of bacterial recombinant MUC1/Z protein,

thereby confirming the identity of the 25 kDa protein as the MUC1/Z protein.

Investigation of the amino acid sequence of the MUC1/Z protein revealed several interesting features.

First, as the MUC1/Z protein contains a signal sequence, but does not harbour a transmembrane domain, it is expected to be secreted from the cell.

Second, an outstanding feature of the MUC1/Z protein is the tryptophan-tryptophan (WW) sequence, localized just proximal to the C-terminal part of the protein [amino acid numbers 93 and 94 in the MUC1/Z sequence (Fig. 8A) and amino acid numbers 102 and 103 in the MUC1/Z/alt sequence (Fig. 8B)]. This is unusual in that tryptophan is the least frequently occurring amino acid in proteins. A computer search for other proteins containing WW sequences revealed that the cell surface receptor for calcitonin contains the sequence G Q R L W W Y H, which is, strikingly, almost identical to the MUC1/Z sequence G Q D L W W Y N [amino acid numbers 89 to 96, Fig. 8A]. Such an occurrence of amino acid identity would occur at a probability of less than 1 in 64 million. This suggests, therefore, that the MUC1/Z protein is in some way involved with cell surface receptor interactions.

Third, the MUC1/Z protein sequence contains several features that are found in other known ligands. For example, human epidermal growth factor (EGF) contains the sequence D L K W W and a similar sequence, D L W W appears in the MUC1/Z protein. Significantly, the location of this sequence is in both proteins identical, and occurs just proximal to the carboxyl-terminus of the protein.

- 44 -

Fourth, a highly-conserved sequence, consisting of C X C X X X X G and which occurs in all growth factor ligand members, appears in the MUC1/Z protein [amino acid numbers 70 to 78, Fig. 8A].

Fifth, the MUC1/Z protein also contains several peptide sequences which are found in members of the prolactin/growth hormone family, such as prolactin, proliferin, and growth hormone.

Taken together, the above considerations all support the present finding that the MUC1/Z protein acts as a ligand for the MUC1/Y receptor protein.

The following experiments further support the above contention. The extracellular domain of the MUC1/Y receptor protein was synthesized as a recombinant bacterial protein and then purified and radioactively labelled, and then was used to probe Western blots containing proteins found in breast tumor tissue lysates. The labelled MUC1/Y receptor protein specifically bound to a 25 kDa protein that comigrated with the MUC1/Z protein; this protein was present in breast tumor tissue lysates, yet was absent in normal breast tissue. Furthermore, in different cell types and tissues, the levels of the MUC1/Z protein directly correlated with the levels of the 25 kDa protein that binds the MUC1/Y receptor protein.

The MUC1/Z protein is therefore the ligand of the MUC1/X, MUC1/Y and MUC1/V receptor proteins. This is true also for MUC1/Z/alt.

MUC1/W and MUC1/W/alt also contain a signal sequence and do not have a transmembrane domain. They are thus

secreted from the cell and, based on their structure, function as ligands in a similar fashion to the MUC1/Z and MUC1/Z/alt proteins.

In the method of the present invention, the new MUC1 proteins described and claimed herein can be administered in various ways. It should be noted that these new MUC1 proteins can be administered alone, or in combination with pharmaceutically acceptable carriers. Compositions according to the present invention can be administered orally or parenterally, including intravenous, intraperitoneal, intranasal and subcutaneous administration. Implants of the compounds are also useful. The patient being treated is a warm-blooded animal, and in particular, mammals including man.

The proteins of the present invention are administered in combination with other drugs, or singly, consistent with good medical practice. The composition is administered and dosed in accordance with good medical practice, taking into account the clinical condition of the individual patient, the site and method of administration, scheduling of administration, and other factors known to medical practitioners. The "effective amount" for purposes herein is thus determined by such considerations as are known in the art.

When administering the new MUC1 proteins parenterally, the pharmaceutical formulations suitable for injection include sterile aqueous solutions or dispersions and sterile powders for reconstitution into sterile injectable solutions or dispersions. The carrier can be a solvent or dispersing medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, liquid polyethylene

- 46 -

glycol, and the like), suitable mixtures thereof, and vegetable oils.

Proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion, and by the use of surfactants. Non-aqueous vehicles such as cottonseed oil, sesame oil, olive oil, soybean oil, corn oil, sunflower oil, or peanut oil and esters, such as isopropyl myristate, may also be used as solvent systems for compound compositions. Additionally, various additives which enhance the stability, sterility, and isotonicity of the compositions, including antimicrobial preservatives, antioxidants, chelating agents, and buffers, can be added. Prevention of the action of microorganisms can be ensured by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, and the like. In many cases, it will be desirable to include isotonic agents, for example, sugars, sodium chloride, and the like. Prolonged absorption of the injectable pharmaceutical form can be brought about by the use of agents delaying absorption, for example, aluminum monostearate and gelatin. According to the present invention, however, any vehicle, diluent or additive used would have to be compatible with the compounds.

Sterile injectable solutions can be prepared by incorporating the proteins utilized in practicing the present invention in the required amount of the appropriate solvent with various of the other ingredients, as desired.

A pharmacological formulation of the new MUC1 proteins described and claimed herein can be administered to the patient in an injectable formulation containing any

- 47 -

compatible carrier, such as various vehicle, adjuvants, additives, and diluents; or the compounds utilized in the present invention can be administered parenterally to the patient in the form of slow-release subcutaneous implants or targeted delivery systems, such as polymer matrices, liposomes, and microspheres. An implant suitable for use in the present invention can take the form of a pellet which slowly dissolves after being implanted, or a biocompatible delivery module well-known to those skilled in the art. Such well-known dosage forms and modules are designed such that the active ingredients are slowly released over a period of several days to several weeks.

Examples of well-known implants and modules useful in the present invention include: U.S. Patent No. 4,487,603, which discloses an implantable micro-infusion pump for dispensing medication at a controlled rate; U.S. Patent No. 4,486,194, which discloses a therapeutic device for administering medicants through the skin; U.S. Patent No. 4,447,233, which discloses a medication infusion pump for delivering medication at a precise infusion rate; U.S. Patent No. 4,447,224, which discloses a variable flow, implantable infusion apparatus for continuous drug delivery; U.S. Patent No. 4,439,196, which discloses an osmotic drug delivery system having multi-chamber compartments; and U.S. Patent No. 4,475,196, which discloses an osmotic drug delivery system. These patents are incorporated herein by reference. Many other such implants, delivery systems, and modules are well-known to those skilled in the art.

A pharmacological formulation of the new MUC1 proteins utilized in the present invention can be administered orally to the patient. Conventional methods such as administering the compounds in tablets, suspensions, solutions, emulsions,

- 48 -

capsules, powders, syrups and the like, are usable. Known techniques which deliver the new MUC1 proteins orally or intravenously and retain the biological activity, are preferred.

In one embodiment, the new MUC1 proteins can be administered initially by intravenous injection to bring blood levels of the new MUC1 proteins to a suitable level. The patient's MUC1 protein levels are then maintained by an oral dosage form, although other forms of administration, dependent upon the patient's condition and as indicated above, can be used. The quantity of the new MUC1 proteins to be administered will vary for the patient being treated, and will vary from about 100 ng/kg of body weight to 100 mg/kg of body weight per day, and preferably will be from 10 µg/kg to 10 mg/kg per day.

EXAMPLE 1

Immunoassays for Detecting and Quantitating the New MUC1 Proteins in Body Fluids

To detect and quantitate the new MUC1 proteins in body fluids such as, for example, serum, one of the most useful methods is the two-antibody sandwich assay [see E. Harlow and D. Lane, *ibid.*, Chapter 14, "Immunoassays," pp. 553-612 (1988)].

Both polyclonal and monoclonal antibodies are prepared against the new MUC1 proteins. To use the two-antibody assay, one antibody is purified and bound to a solid phase, and one of the new MUC1 proteins which is to be assayed is allowed to bind. Unbound proteins are removed by washing and the labelled second antibody is allowed to bind to the

- 49 -

antigen. After washing, the assay is quantitated by measuring the amount of labelled second antibody that is bound to the matrix and a calibration curve is established for the specific new MUC1 protein which was assayed.

To assay for the presence of the new MUC1 proteins in body fluids, the above assay is repeated, using as test antigen a sample of the body fluid.

EXAMPLE 2

Immunohistochemical Staining for the Detection of the New MUC1 Proteins in Tissue Sections

Histological studies for the detection of the new MUC1 proteins are carried out on paraformaldehyde-fixed, paraffin-embedded tissue samples.

The cells or tissues are fixed to the glass slides and permeabilized using standard procedures as described in E. Harlow and D. Lane, *ibid.*, Chapter 10, "Cell Staining," pp. 359-420 (1988). The antibodies against one of the new MUC1 proteins are then added to the fixed and permeabilized cells or tissues. As in many other immuno-chemical techniques, the antibodies can be labelled directly either with an enzyme, fluorochrome, etc., or detected by using a labelled secondary reagent that binds specifically to the primary antibody.

EXAMPLE 3In-Vivo Imaging of Breast Cancer Cells with Labelled Ligands that Bind to the New MUC1 Receptor Proteins

The MUC1/Z, MUC1/Z/alt, MUC1/W and MUC1/W/alt ligand proteins are used to target and thereby image breast cancer cells in the living body. These ligand molecules are radioactively labelled with, for example, radioactive iodine (^{125}I) using, for example, the Bolton-Hunter reagent [^{125}I -labelled N-succinimidyl 3-(4-hydroxy-phenylpropionate)].

An 0.5-1 mg/ml solution of the new MUC12 ligand proteins is prepared in 0.1 M sodium borate (pH 8.5) and transferred to ice. Approximately 500 microcurie of Bolton-Hunter reagent is transferred to a 1.5 ml conical tube at 0°C and the reagent is dried in a stream of dry nitrogen gas. About 10 microliters of the protein solution is added to the dry Bolton-Hunter reagent, mixed gently and returned to the ice. Following incubation on ice for 15 minutes, a stop solution consisting of 100 microliters of 0.5 M ethanolamine, 10% glycerol, 0.1% xylene cyanol, 0.1 M sodium borate (pH 8.5) is added and incubated for 5 min. at room temperature. The radioactively iodinated MUC1/Z, MUC1/Z/alt, MUC1/W and MUC1/W/alt ligand proteins are then separated from the iodinated Bolton-Hunter reagent on a gel-filtration column.

To image breast cancer cells in vivo, the labelled ligand molecules are injected intravenously into the patient, and the distribution of the radioactively labelled molecules is monitored using radioactive imaging devices.

- 51 -

EXAMPLE 4Ligand as a Drug Delivery System for Ligand-Toxin Conjugates

The MUC1/Z, MUC1/Z/alt, MUC1/W and MUC1/W/alt ligand proteins are conjugated to cytotoxic substances and thereby used as drug delivery systems to target and kill breast cancer cells within the body. Several cytotoxic substances for conjugation may be used, including cytotoxic proteins such as pseudomonas exotoxin A and ricin [I. Pastan and D. Fitzgerald, "Recombinant Toxins for Cancer Treatment," Science, Vol. 254, pp. 1173-1177 (1991)] or cytotoxic levels of radioactivity.

Conjugation of the new MUC1 proteins to cytotoxic proteins is performed by any of a number of coupling procedures, including glutaraldehyde coupling and periodate coupling.

In the two-step glutaraldehyde method, glutaraldehyde is first coupled to the pure cytotoxic protein via the reactive amino groups available on the protein. The cytotoxic protein-glutaraldehyde mix is then purified and added to the MUC1/Z, MUC1/Z/alt, MUC1/W, and MUC1/W/alt ligand proteins. Unconjugated material is then separated from the cytotoxic protein/new MUC1 protein conjugate.

The cytotoxic protein is dissolved in 0.2 ml of 1.25% glutaraldehyde (electron microscopic grade) in 100 mM sodium phosphate (pH 6.8). After 18 hours at room temperature, excess free glutaraldehyde is removed by gel filtration on a gel matrix that is pre-equilibrated with 0.15 M NaCl. The peak fractions containing the glutaraldehyde-linked cytotoxic protein are concentrated by ultrafiltration or by dialysis against 100 mM sodium carbonate-sodium bicarbonate

buffer (pH 9.5) containing 30% sucrose. The new MUC1 ligand proteins dissolved in 0.1 ml of 0.15 M NaCl are added to the cytotoxic protein solution, the pH is kept above 9.0, and the mixture is incubated at 4°C for 24 hours. At this stage, 0.1 ml of 0.2 M ethanolamine (pH 7.0) is added and the mixture incubated for a further 2 hours at 4°C.

The cytotoxic protein-new MUC1 ligand conjugate is then separated from the unconjugated protein molecules by either gel filtration or gel electrophoresis.

For periodate coupling, the new MUC1 ligand proteins are resuspended in 1.2 ml of water and freshly-prepared 0.1 M sodium periodate (0.3 ml) in 10 mM sodium phosphate buffer (pH 7.0) is added. The mixture is incubated at room temperature for 20 minutes and then dialysed against 1 mM sodium acetate (pH 4.0) at 4°C with several changes overnight. A 0.5 ml solution (10 mg/ml) of the cytotoxic protein (for example, ricin) is prepared in 20 mM sodium carbonate buffer (pH 9.5) and added to the solution of the periodate treated new MUC1 ligand proteins. The mixture is incubated at room temperature for 2 hours. The Schiff's bases that have formed are then reduced by adding 100 microliters of sodium borohydride (4 mg/ml) in water and incubating at 4°C for 2 hours.

The cytotoxic protein-new MUC1 ligand conjugate is then separated from the unconjugated protein molecules by either gel filtration or gel electrophoresis.

Cytotoxic protein-new MUC1 ligand conjugates may also be prepared using recombinant DNA technology. In this method, recombinant bacteria are generated that synthesize

- 53 -

fusion proteins consisting of the cytotoxic protein fused to the new MUC1 ligand proteins.

It will be evident to those skilled in the art that the invention is not limited to the details of the foregoing illustrative embodiments and examples, and that the present invention may be embodied in other specific forms without departing from the essential attributes thereof, and it is therefore desired that the present embodiments be considered in all respects as illustrative and not restrictive, reference being made to the appended claims, rather than to the foregoing description, and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein.

WHAT IS CLAIMED IS:

1. A biochemically pure MUC1 protein, selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, MUC1/W, MUC1/W/alt, MUC1/Z and MUC1/Z/alt or a functional derivative thereof devoid of a tandem repeat array.

2. A MUC1 protein according to claim 1 or a functional derivative thereof, comprising a partial amino acid sequence

M T P G T Q S P F F L L L L L T V L T [A T T A P K P A T]
V V T G S G H A S S T P G G E K E T S A T Q R S S V P
S S T E K N A

and devoid of a tandem repeat array downstream thereof.

3. A MUC1 protein according to claim 1 or a functional derivative thereof, comprising a partial amino acid sequence

M T P G T Q S P F F L L L L L T V L T [A T T A P K P A T]
V V T G S G H A S S T P G G E K E T S A T Q R S S V P S

and devoid of a tandem repeat array downstream thereof.

4. Biochemically pure MUC1/X and MUC1/X/alt, respectively comprising the sequences shown in Figs. 5A and 5B, or functional derivatives thereof.

- 55 -

5. Biochemically pure MUC1/Y and MUC1/Y/alt, respectively comprising the sequences shown in Figs. 6A and 6B, or functional derivatives thereof.

6. Biochemically pure MUC1/V and MUC1/V/alt, respectively comprising the sequences shown in Figs. 6C and 6D, or functional derivatives thereof.

7. Biochemically pure MUC1/W and MUC1/W/alt, respectively comprising the sequences shown in Figs. 7A and 7B, or functional derivatives thereof.

8. Biochemically pure MUC1/Z and MUC1/Z/alt, respectively comprising the sequences shown in Figs. 8A and 8B, or functional derivatives thereof.

9. A pharmaceutical composition, comprising as an active ingredient therein a biochemically pure MUC1 protein selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, MUC1/W, MUC1/W/alt, MUC1/Z, MUC1/Z/alt and functional derivatives thereof, in combination with a pharmaceutically acceptable carrier.

10. A pharmaceutical composition for the treatment of human breast cancer, comprising as an active ingredient therein a biochemically pure MUC1 protein selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, and functional derivatives thereof, in soluble form and in combination with a pharmaceutically acceptable carrier.

11. A conjugated toxin for the treatment of human breast cancer, comprising a MUC1 protein selected from the group consisting of MUC1/W, MUC1/W/alt, MUC1/Z, MUC1/Z/alt and functional derivatives thereof, attached to a cytotoxic agent.

12. A diagnostic agent for the detection of human breast cancer cells, comprising a detectably labelled, biochemically pure MUC1 protein selected from the group consisting of MUC1/W, MUC1/W/alt, MUC1/Z, MUC1/Z/alt and functional derivatives thereof.

13. A diagnostic agent for identification of sites in the body to which breast cancer cells have spread, comprising a detectably labelled MUC1 protein selected from the group consisting of MUC1/W, MUC1/W/alt, MUC1/Z, MUC1/Z/alt and functional derivatives thereof.

- 57 -

14. A method for the treatment of human breast cancer, comprising administering to an individual having human breast cancer cells an amount of soluble MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, or MUC1/V/alt receptors, sufficient to inhibit the binding of MUC1 ligands to said cells.

15. A method for the treatment of human breast cancer, comprising administering to an individual having human breast cancer cells an amount of a ligand-toxin conjugant comprising a ligand selected from MUC1/W, MUC1/W/alt, MUC1/Z or MUC1/Z/alt, fused to a cytotoxic toxin.

16. A DNA sequence encoding the protein MUC1/X, comprising the nucleotide sequence substantially as shown in Fig. 5A or a functional derivative thereof devoid of a tandem repeat array.

17. A DNA sequence encoding the protein MUC1/X/alt, comprising the nucleotide sequence substantially as shown in Fig. 5B or a functional derivative thereof devoid of a tandem repeat array.

18. A DNA sequence encoding the protein MUC1/Y, comprising the nucleotide sequence substantially as shown in Fig. 6A or a functional derivative thereof devoid of a tandem repeat array.

19. A DNA sequence encoding the protein MUC1/Y/alt, comprising the nucleotide sequence substantially as shown in Fig. 6B or a functional derivative thereof devoid of a tandem repeat array.

20. A DNA sequence encoding the protein MUC1/V, comprising the nucleotide sequence substantially as shown in Fig. 6C or a functional derivative thereof devoid of a tandem repeat array.

21. A DNA sequence encoding the protein MUC1/V/alt, comprising the nucleotide sequence substantially as shown in Fig. 6D or a functional derivative thereof devoid of a tandem repeat array.

22. A DNA sequence encoding the protein MUC1/W, comprising the nucleotide sequence substantially as shown in Fig. 7A or a functional derivative thereof devoid of a tandem repeat array.

23. A DNA sequence encoding the protein MUC1/W/alt, comprising the nucleotide sequence substantially as shown in Fig. 7B or a functional derivative thereof devoid of a tandem repeat array.

24. A DNA sequence encoding the protein MUC1/Z, comprising the nucleotide sequence substantially as shown in Fig. 8A or a functional derivative thereof devoid of a tandem repeat array.

25. A DNA sequence encoding the protein MUC1/Z/alt, comprising the nucleotide sequence substantially as shown in Fig. 8B or a functional derivative thereof devoid of a tandem repeat array.

26. A DNA sequence according to any of claims 16-25, being a cDNA.

27. An in-vitro bioassay for determining the presence of breast cancer cells in a sample, comprising contacting a tissue sample with a diagnostic agent, said agent comprising a detectable labelled MUC1 protein selected from the group consisting of MUC1/W, MUC1/W/alt, MUC1/Z, MUC1/Z/alt and functional derivatives thereof.

28. An in-vitro bioassay for determining the presence of breast cancer cells in a sample, comprising:

a) isolating a specimen selected from the group consisting of tissue and cell biopsies, and

b) assaying said specimen with antibodies selected from the group consisting of monoclonal and polyclonal antibodies that recognize a protein selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, MUC1/W, MUC1/W/alt, MUC1/Z, MUC1/Z/alt and functional derivatives thereof.

29. A DNA construct selected from the group consisting of cDNA coding for a biochemically pure MUC1 protein selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, MUC1/W, MUC1/W/alt, MUC1/Z and MUC1/Z/alt or a functional derivative thereof devoid of a tandem repeat array.

30. The construct of claim 29, which is contained in a vector.

31. A host cell transfected with the construct of claim 30.

32. A bioassay for screening substances for the ability to inhibit mammary carcinoma, comprising:

a) administering the substance to a cell transfectant that expresses a protein selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, and functional derivatives thereof; and

b) determining whether such substance inhibits the growth of the cell transfectant.

33. A purified antibody which specifically binds a protein of claim 1.

34. The antibody of claim 33, wherein said antibody is conjugated to a therapeutic drug.

- 61 -

35. The antibody of claim 33, wherein said antibody is conjugated to a detectable moiety.

36. The antibody of claim 33, wherein said antibody is bound to a solid support.

37. A bioassay for determining the amount of a MUC1 protein selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, MUC1/W, MUC1/W/alt, MUC1/Z and MUC1/Z/alt or a functional derivative thereof devoid of a tandem repeat array in a biological sample, comprising:

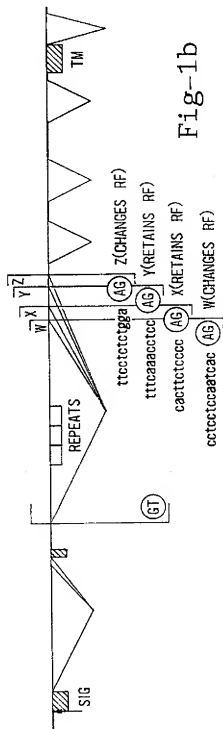
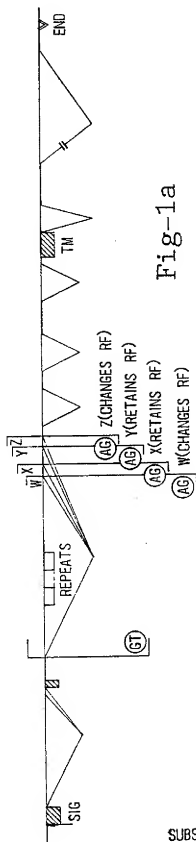
a) contacting said biological sample with an antibody under conditions such that a specific complex of said antibody and said MUC1 protein can be formed; and

b) determining the amount of said antibody/MUC1 protein complex, the amount of the complex indicating the amount of said MUC1 protein in the biological sample.

38. A method of detecting the presence of cancer in a subject, comprising determining the presence of a detectable amount of a MUC1 protein selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, MUC1/W, MUC1/W/alt, MUC1/Z and MUC1/Z/alt or a functional derivative thereof devoid of a tandem repeat array in a biopsy from said subject, the presence of a detectable amount of said MUC1 protein relative to the absence of said MUC1 protein in a normal control indicating the presence of cancer.

39. A method of determining the prognosis of a subject having cancer, comprising determining the presence of a detectable amount of a MUC1 protein selected from the group consisting of MUC1/X, MUC1/X/alt, MUC1/Y, MUC1/Y/alt, MUC1/V, MUC1/V/alt, MUC1/W, MUC1/W/alt, MUC1/Z and MUC1/Z/alt or a functional derivative thereof devoid of a tandem repeat array in a biopsy from in said subject, the presence of a detectable amount of said MUC1 protein relative to the absence of said MUC1 protein in a normal control indicating a decreased chance of long-term survival.

1/15



2/15

72 81 104 143 175 211 288 345 390

M T P G T Q S P F F L L L L L T V L T V V
 T G S G H A S S T P G G E K E T S A T Q
 375 463 373 310 276 260 114 159 113 51 101 93 79 106 82 60 47 129 53



72 81 104 143 175 217 306 425 395

M T P G T Q S P F F L L L L L T V L T A T
T A P K P A T V V T G S G H A S S T P G
 452 521 421 327 273 288 139 170 146 81 95 55 88 95 129 56 106 58 51 101



G E K E T S A T Q
 93 79 106 82 60 47 129 53

Fig-2

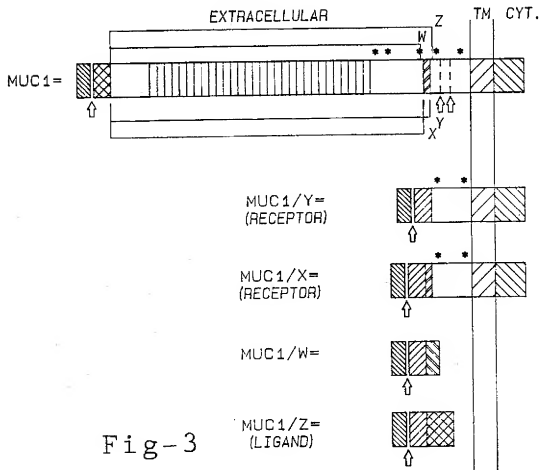


Fig-3

3/15

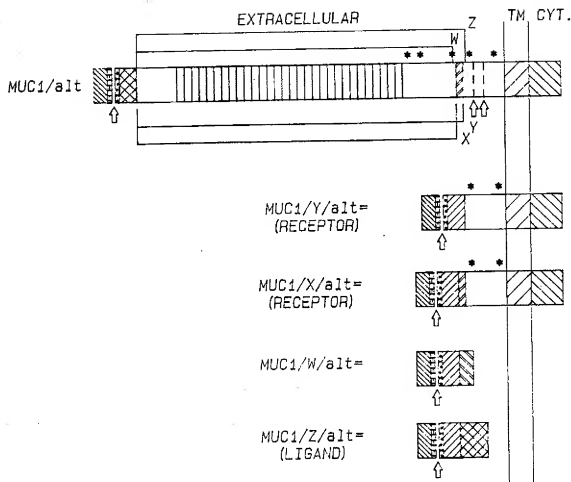


Fig-4

4/15

30
 ATGACACCGGGACCCAGTCTCCTTTCTTCTGCTGCTCCTCACAGTGTCTTACAGTT
 M T P G T Q S P F F L L L L L T V L T V 60
 90
 GTTACAGGTTCTGGTCATGCAAGCTCTACCCAGGTGGAGAAAAGGAGACTTCGGGTACC
 V T G S G H A S S T P G G E K E T S A T 120
 150
 CAGAGAAGTTCAAGTCCCCAGCTCTACTGAGAAGAATGCTTTGTCTACTGGGGTCTCTTTC
 Q R S S V P S S T E K N A L S T G V S F 180
 210
 TTTTCTGCTCTTTTTCACATTTCAAACCTCCAGTTTAACTCTCTGGAGAGATCCCAGC
 F F L S F H I S N L G F N S S L E D P S 240
 270
 ACCGACTACTACCAAGAGCTGCAGAGAGACATTTCTGAAATGTTTTGCAGATTTATAA
 T D Y Y G E L Q R D I S E M F L Q I Y K 300
 330
 CAAGGGGGTTTCTGGGCTCTCCAATATTAAGTTCAAGGCGAGGATCTGTGGTGGTACAA
 Q G G G F L G L S N I K F R P G S V V V Q 360
 390
 TTGACTCTGGCCTTCGAGAGGTTACCATCAATGTCCACGAGTGGAGACACAGTTCAAT
 L T L A F R E G T I N V H D V E T Q F N 420
 450
 CAGTATAAAACGGAAGCAGCTCTCGATATAAACCTGACGATCTCAGACGTCAGCGTGAGT
 G Y K T E A A S R Y N L T I S D V S V S 480
 510
 GATGTGCCATTTCTTTCTCTGCCCAGTCTGGGGCTGGGGTGCCAGGCTGGGGCATCGG
 D V P F P F S A G S G A G V P G W G I A 540
 570
 CTGCTGGTGTGGTCTGTGTCTGCTGGCTGGCCATTGTCTATCTCAITGCTTGGCT
 L L V L V C V L V A L A I V Y L I A L A 600
 630
 GTCTGTCACTGCCCGGAAAGAAGTACGGGCACTGGACATCTTTCCAGCCCGGATACC
 V C G C R R K N Y G Q L D I F P A R D T 660
 690
 TACCATCCTATGAGCGAGTACCCCACTACCAACCCATGGGCGCTATGTGCCCTTACG
 Y H P M S E Y P T Y H T H G R Y V P P S 720
 750
 AGTACCGATCTGTAGCCCTATGAGAAGGTTTCTGCAAGTAAATGGTGGCAGCAGCCTCTCT
 S T D R S P Y E K V S A G N G G S S L S 780
 810
 TACACAAACCCAGCAGTGGCAGCCACTTGTGCAACTTGTAG
 Y T N P A V A A T S A N L U 840

Fig-5A

5/15
 30
 60
 120
 180
 240
 300
 360
 420
 480
 540
 600
 660
 720
 780
 840
 900
 960
 1020
 1080
 1140
 1200
 1260
 1320
 1380
 1440
 1500
 1560
 1620
 1680
 1740
 1800
 1860
 1920
 1980
 2040
 2100
 2160
 2220
 2280
 2340
 2400
 2460
 2520
 2580
 2640
 2700
 2760
 2820
 2880
 2940
 3000
 3060
 3120
 3180
 3240
 3300
 3360
 3420
 3480
 3540
 3600
 3660
 3720
 3780
 3840
 3900
 3960
 4020
 4080
 4140
 4200
 4260
 4320
 4380
 4440
 4500
 4560
 4620
 4680
 4740
 4800
 4860
 4920
 4980
 5040
 5100
 5160
 5220
 5280
 5340
 5400
 5460
 5520
 5580
 5640
 5700
 5760
 5820
 5880
 5940
 6000
 6060
 6120
 6180
 6240
 6300
 6360
 6420
 6480
 6540
 6600
 6660
 6720
 6780
 6840
 6900
 6960
 7020
 7080
 7140
 7200
 7260
 7320
 7380
 7440
 7500
 7560
 7620
 7680
 7740
 7800
 7860
 7920
 7980
 8040
 8100
 8160
 8220
 8280
 8340
 8400
 8460
 8520
 8580
 8640
 8700
 8760
 8820
 8880
 8940
 9000
 9060
 9120
 9180
 9240
 9300
 9360
 9420
 9480
 9540
 9600
 9660
 9720
 9780
 9840
 9900
 9960
 10020
 10080
 10140
 10200
 10260
 10320
 10380
 10440
 10500
 10560
 10620
 10680
 10740
 10800
 10860
 10920
 10980
 11040
 11100
 11160
 11220
 11280
 11340
 11400
 11460
 11520
 11580
 11640
 11700
 11760
 11820
 11880
 11940
 12000
 12060
 12120
 12180
 12240
 12300
 12360
 12420
 12480
 12540
 12600
 12660
 12720
 12780
 12840
 12900
 12960
 13020
 13080
 13140
 13200
 13260
 13320
 13380
 13440
 13500
 13560
 13620
 13680
 13740
 13800
 13860
 13920
 13980
 14040
 14100
 14160
 14220
 14280
 14340
 14400
 14460
 14520
 14580
 14640
 14700
 14760
 14820
 14880
 14940
 15000
 15060
 15120
 15180
 15240
 15300
 15360
 15420
 15480
 15540
 15600
 15660
 15720
 15780
 15840
 15900
 15960
 16020
 16080
 16140
 16200
 16260
 16320
 16380
 16440
 16500
 16560
 16620
 16680
 16740
 16800
 16860
 16920
 16980
 17040
 17100
 17160
 17220
 17280
 17340
 17400
 17460
 17520
 17580
 17640
 17700
 17760
 17820
 17880
 17940
 18000
 18060
 18120
 18180
 18240
 18300
 18360
 18420
 18480
 18540
 18600
 18660
 18720
 18780
 18840
 18900
 18960
 19020
 19080
 19140
 19200
 19260
 19320
 19380
 19440
 19500
 19560
 19620
 19680
 19740
 19800
 19860
 19920
 19980
 20040
 20100
 20160
 20220
 20280
 20340
 20400
 20460
 20520
 20580
 20640
 20700
 20760
 20820
 20880
 20940
 21000
 21060
 21120
 21180
 21240
 21300
 21360
 21420
 21480
 21540
 21600
 21660
 21720
 21780
 21840
 21900
 21960
 22020
 22080
 22140
 22200
 22260
 22320
 22380
 22440
 22500
 22560
 22620
 22680
 22740
 22800
 22860
 22920
 22980
 23040
 23100
 23160
 23220
 23280
 23340
 23400
 23460
 23520
 23580
 23640
 23700
 23760
 23820
 23880
 23940
 24000
 24060
 24120
 24180
 24240
 24300
 24360
 24420
 24480
 24540
 24600
 24660
 24720
 24780
 24840
 24900
 24960
 25020
 25080
 25140
 25200
 25260
 25320
 25380
 25440
 25500
 25560
 25620
 25680
 25740
 25800
 25860
 25920
 25980
 26040
 26100
 26160
 26220
 26280
 26340
 26400
 26460
 26520
 26580
 26640
 26700
 26760
 26820
 26880
 26940
 27000
 27060
 27120
 27180
 27240
 27300
 27360
 27420
 27480
 27540
 27600
 27660
 27720
 27780
 27840
 27900
 27960
 28020
 28080
 28140
 28200
 28260
 28320
 28380
 28440
 28500
 28560
 28620
 28680
 28740
 28800
 28860
 28920
 28980
 29040
 29100
 29160
 29220
 29280
 29340
 29400
 29460
 29520
 29580
 29640
 29700
 29760
 29820
 29880
 29940
 30000
 30060
 30120
 30180
 30240
 30300
 30360
 30420
 30480
 30540
 30600
 30660
 30720
 30780
 30840
 30900
 30960
 31020
 31080
 31140
 31200
 31260
 31320
 31380
 31440
 31500
 31560
 31620
 31680
 31740
 31800
 31860
 31920
 31980
 32040
 32100
 32160
 32220
 32280
 32340
 32400
 32460
 32520
 32580
 32640
 32700
 32760
 32820
 32880
 32940
 33000
 33060
 33120
 33180
 33240
 33300
 33360
 33420
 33480
 33540
 33600
 33660
 33720
 33780
 33840
 33900
 33960
 34020
 34080
 34140
 34200
 34260
 34320
 34380
 34440
 34500
 34560
 34620
 34680
 34740
 34800
 34860
 34920
 34980
 35040
 35100
 35160
 35220
 35280
 35340
 35400
 35460
 35520
 35580
 35640
 35700
 35760
 35820
 35880
 35940
 36000
 36060
 36120
 36180
 36240
 36300
 36360
 36420
 36480
 36540
 36600
 36660
 36720
 36780
 36840
 36900
 36960
 37020
 37080
 37140
 37200
 37260
 37320
 37380
 37440
 37500
 37560
 37620
 37680
 37740
 37800
 37860
 37920
 37980
 38040
 38100
 38160
 38220
 38280
 38340
 38400
 38460
 38520
 38580
 38640
 38700
 38760
 38820
 38880
 38940
 39000
 39060
 39120
 39180
 39240
 39300
 39360
 39420
 39480
 39540
 39600
 39660
 39720
 39780
 39840
 39900
 39960
 40020
 40080
 40140
 40200
 40260
 40320
 40380
 40440
 40500
 40560
 40620
 40680
 40740
 40800
 40860
 40920
 40980
 41040
 41100
 41160
 41220
 41280
 41340
 41400
 41460
 41520
 41580
 41640
 41700
 41760
 41820
 41880
 41940
 42000
 42060
 42120
 42180
 42240
 42300
 42360
 42420
 42480
 42540
 42600
 42660
 42720
 42780
 42840
 42900
 42960
 43020
 43080
 43140
 43200
 43260
 43320
 43380
 43440
 43500
 43560
 43620
 43680
 43740
 43800
 43860
 43920
 43980
 44040
 44100
 44160
 44220
 44280
 44340
 44400
 44460
 44520
 44580
 44640
 44700
 44760
 44820
 44880
 44940
 45000
 45060
 45120
 45180
 45240
 45300
 45360
 45420
 45480
 45540
 45600
 45660
 45720
 45780
 45840
 45900
 45960
 46020
 46080
 46140
 46200
 46260
 46320
 46380
 46440
 46500
 46560
 46620
 46680
 46740
 46800
 46860
 46920
 46980
 47040
 47100
 47160
 47220
 47280
 47340
 47400
 47460
 47520
 47580
 47640
 47700
 47760
 47820
 47880
 47940
 48000
 48060
 48120
 48180
 48240
 48300
 48360
 48420
 48480
 48540
 48600
 48660
 48720
 48780
 48840
 48900
 48960
 49020
 49080
 49140
 49200
 49260
 49320
 49380
 49440
 49500
 49560
 49620
 49680
 49740
 49800
 49860
 49920
 49980
 50040
 50100
 50160
 50220
 50280
 50340
 50400
 50460
 50520
 50580
 50640
 50700
 50760
 50820
 50880
 50940
 51000
 51060
 51120
 51180
 51240
 51300
 51360
 51420
 51480
 51540
 51600
 51660
 51720
 51780
 51840
 51900
 51960
 52020
 52080
 52140
 52200
 52260
 52320
 52380
 52440
 52500
 52560
 52620
 52680
 52740
 52800
 52860
 52920
 52980
 53040
 53100
 53160
 53220
 53280
 53340
 53400
 53460
 53520
 53580
 53640
 53700
 53760
 53820
 53880
 53940
 54000
 54060
 54120
 54180
 54240
 54300
 54360
 54420
 54480
 54540
 54600
 54660
 54720
 54780
 54840
 54900
 54960
 55020
 55080
 55140
 55200
 55260
 55320
 55380
 55440
 55500
 55560
 55620
 55680
 55740
 55800
 55860
 55920
 55980
 56040
 56100
 56160
 56220
 56280
 56340
 56400
 56460
 56520
 56580
 56640
 56700
 56760
 56820
 56880
 56940
 57000
 57060
 57120
 57180
 57240
 57300
 57360
 57420
 57480
 57540
 57600
 57660
 57720
 57780
 57840
 57900
 57960
 58020
 58080
 58140
 58200
 58260
 58320
 58380
 58440
 58500
 58560
 58620
 58680
 58740
 58800
 58860
 58920
 58980
 59040
 59100
 59160
 59220
 59280
 59340
 59400
 59460
 59520
 59580
 59640
 59700
 59760
 59820
 59880
 59940
 60000
 60060
 60120
 60180
 60240
 60300
 60360
 60420
 60480
 60540
 60600
 60660
 60720
 60780
 60840
 60900
 60960
 61020
 61080
 61140
 61200
 61260
 61320
 61380
 61440
 61500
 61560
 61620
 61680
 61740
 61800
 61860
 61920
 61980
 62040
 62100
 62160
 62220
 62280
 62340
 62400
 62460
 62520
 62580
 62640
 62700
 62760
 62820
 62880
 62940
 63000
 63060
 63120
 63180
 63240
 63300
 63360
 63420
 63480
 63540
 63600
 63660
 63720
 63780
 63840
 63900
 63960
 64020
 64080
 64140
 64200
 64260
 64320
 64380
 64440
 64500
 64560
 64620
 64680
 64740
 64800
 64860
 64920
 64980
 65040
 65100
 65160
 65220
 65280
 65340
 65400
 65460
 65520
 65580
 65640
 65700
 65760
 65820
 65880
 65940
 66000
 66060
 66120
 66180
 66240
 66300
 66360
 66420
 66480
 66540
 66600
 66660
 66720
 66780
 66840
 66900
 66960
 67020
 67080
 67140
 67200
 67260
 67320
 67380
 67440
 67500
 67560
 67620
 67680
 67740
 67800
 67860
 67920
 67980
 68040
 68100
 68160
 68220
 68280
 68340
 68400
 68460
 68520
 68580
 68640
 68700
 68760
 68820
 68880
 68940
 69000
 69060
 69120
 69180
 69240
 69300
 69360
 69420
 69480
 69540
 69600
 69660
 69720
 69780
 69840
 69900
 69960
 70020
 70080
 70140
 70200
 70260
 70320
 70380
 70440
 70500
 70560
 70620
 70680
 70740
 70800
 70860
 70920
 70980
 71040
 71100
 71160
 71220
 71280
 71340
 71400
 71460
 71520
 71580
 71640
 71700
 71760
 71820
 71880
 71940
 72000
 72060
 72120
 72180
 72240
 72300
 72360
 72420
 72480
 72540
 72600
 72660
 72720
 72780
 72840
 72900
 72960
 73020
 73080
 73140
 73200
 73260
 73320
 73380
 73440
 73500
 73560
 73620
 73680
 73740
 73800
 73860
 73920
 73980
 74040
 74100
 74160
 74220
 74280
 74340
 74400
 74460
 74520
 74580
 74640

6/15

30
 ATGACACCGGGCACCAGTCTCCTTTCTTCTGCTGCTGCTCCTCACAGTGTCTACAGTT
 M T P G T Q S P F F L L L L L T V L T V
 60
 90
 GTTACAGGTTCTGGTCATGCAAGCTCTACCCAGGTGGAGAAAAGGAGACTTCGGCTACC
 V T G S G H A S S T P G G E K E T S A T
 120
 150
 CAGAGAAGTTCAGTGCCCACTCTACTGAGAAGAAATGCTTTTAATTCCTCTCTGGAAGAT
 Q R S S V P S S T E K N A F N S S L E D
 180
 210
 CCCAGCACCAGTACTACTACCAAGAGCTGCAAGAGACATTCTGAAATGTTTTGCAGATT
 P S T D Y Y G E L G R D I S E M F L G I
 240
 270
 TATAACAAGGGGTTTTCTGGGCTCTCCAATATTAAGTTCAGGCCAGGATCTGTGGTG
 Y K G G G F L G L S N I K F R P G S V V
 300
 330
 GTACAATTGACTCTGGCTTCCGAGAAGGTACCATCAATGTCCACGACGTGGAGACACAG
 V G L T L A F R E G T I N V H D V E T Q
 360
 390
 TTCAATCAGTATAAAACGGGAAGCAGCCTCTCGATATAACCTGACGATCTCAGACGTACG
 F N Q Y K T E A A S R Y N L T I S D V S
 420
 450
 GTGAGTGATGTGCCATTTCCTTTCTCTGCCAGTCTGGGGCTGGGGTGCCAGGCTGGGGC
 V S D V P F P F S A Q S G A G V P G W G
 480
 510
 ATCGCGCTGCTGGTCTGGTCTGTGTTCTGGTGGCTGGGCAATGCTCTATCTCATTTGCC
 I A L L V L V C V L V A L A I V Y L I A
 540
 570
 TTGGCTGTCTGTCAGTGCCGCCGAAAGAACTACGGGCGAGCTGGACATCTTCCAGCCCGG
 L A V C Q C R R K N Y G Q L D I F P A H
 600
 630
 GATACCTACCATCCTATGAGCGAGTACCCCACTACCAACCCATGGGCGCTATGTGCC
 D T Y H P M S E Y P T Y H T H G R Y V P
 660
 690
 CCTAGCAGTACCGATCGTAGCCCTATGAGAAGGTTTCTGCAGGTAATGTTGGCAGCAGC
 P S S T D R S P Y E K V S A G N G G S S
 720
 750
 CTCTCTTACACAAACCCAGCAGTGGCAGCCACTTCTGCCAAGTTGTAG
 L S Y T N P A V A A T S A N L U
 780

Fig-6A

7/15

ATGACACCGGGCACCCAGTCTCCTTTCT³⁰CTTCTGCTGCTGCTCCTCACAGTGTCTTACAGTT⁶⁰
 M T P G T G S P F F L L L L L T V L T V
 ACCACAGCCCTAAACCCGCAACAGTTGTTACAGGTTCTG⁹⁰GTATGCAAGCTCTACCCCA¹²⁰
 T T A P K P A T V V T G S G H A S S T P
 GGTTGGAGAAAAGGAGACTTCGGCTACCCAGAGAAAGTTCA¹⁵⁰GTGCCAGCTCTACTGAGAAG¹⁸⁰
 G G E K E T S A T G H S S V P S S T E K
 AATGCTTTTAAATTCCTCTCTGGAAGATCCCAGCACCAGCT²¹⁰ACTACCAAGAGCTGCAGAGA²⁴⁰
 N A F N S S L E D P S T D Y Y Q E L G H
 GACATTTCTGAAATGTTTTGCAGATTATAAAACAAGGGGG²⁷⁰TTTCTGGGCTCTCCAAT³⁰⁰
 D I S E M F L Q I Y K Q G G F L G L S N
 ATTAAGTTCAGGCCAGGATCTGTGGTGGTACAATTGACTCT³³⁰GGCCTTCGGAAGGTACC³⁶⁰
 I K F H P G S V V V Q L T L A F H E G T
 ATCAATGTCCACGACGTGGAGACACAGTTCAATCAGTATA³⁹⁰AAACGGAAGCAGCCTCTCGA⁴²⁰
 I N V H D V E T Q F N Q Y K T E A A S H
 TATAACCTGACGATCTCAGACGTCAGCGTGAGTGATGTG⁴⁵⁰CCATTCTCTGTGCCAG⁴⁸⁰
 Y N L T I S D V S V S D V P F P F S A Q
 TCTGGGGCTGGGGTGCCAGGGCTGCGGCTGCTGGTGTG⁵¹⁰CTGTGTGTTCTGTGTT⁵⁴⁰
 S G A G V P G W G I A L L V L V C V L V
 GCGCTGGCCATTGTCTATCTCAITGCTTGGCTGTCTGT⁵⁷⁰CAGTGCCGCCGAAAGAATAC⁶⁰⁰
 A L A I V Y L I A L A V C Q C R R K N Y
 GGGCAGCTGGACATCTTTCCAGCCCGGGATACCTACCAT⁶³⁰CCTATGACCGAGTACCCCA⁶⁶⁰
 G Q L D I F P A R D T Y H P M S E Y P T
 TACCACACCATGGGCGCTATGTGCCCGCTAGCAGTACC⁶⁹⁰GATCGTAGCCCTATGAGAAG⁷²⁰
 Y H T H G R Y V P P S S T D R S P Y E K
 GTTCTGCAAGGTAATGGTGGCAGCAGCCTCTCTTACACA⁷⁵⁰AAACCCAGCAGTGGCAGCCACT⁷⁸⁰
 V S A G N G G S S L S Y T N P A V A A T
 TCTGCCAACITGTAG²⁶⁰
 S A N L U

Fig-6B

9/15

30 60
 ATGACACCGGGCACCCAGTCTCCTTTCTTCCTGCTGCTGCTCCTCAGTGTCTACAGCT
 M T P G T Q S P F F L L L L L T V L T A 20
 90 120
 ACCACAGCCCTTAAACCCGCAACAGTTGTTACAGGTTCTGGTCATGCAAGCTCTACCCCA
 T T A P K P A T V V T G S G H A S S T P 40
 150 180
 GGTGGAGAAAAGGAGACTTCGGCTACCCAGAGAAAGTTCAAGTCCAGCAGCAGGACTACTAC
 G G E K E T S A T Q R S S V P S T D Y V 60
 210 240
 CAAGAGCTGCAGAGAGACATTTCTGAAATGTTTTGCAATTTATAAACAAAGGGGGTTTT
 G E L G R D I S E M F L Q I Y K G G G F 80
 270 300
 CTGGGCTCTCCAATATTAAGTTACAGGCCAGGATCTGTGTGGTACAATGACTCTGCC
 L G L S N I K F R P G S V V V G L T L A 100
 330 360
 TTCCGAGAAAGTACCATCAATGTCCACGACGTGGAGACACAGTTCAATCAGTATAAAACG
 F H E G T I N V H D V E T Q F N G Y K T 120
 390 420
 GAAAGACCTCTCGATATAACCTGACGATCTCAGACGTCAGCGTGAGTGATGTGCCATTT
 E A A S R Y N L T I S D V S V S D V P F 140
 450 480
 CTTTCTCTGCCAGTCTGGGGCTGGGGTGCCAGGCTGGGGCATCGCGCTGCTGGTGCTG
 P F S A G S G A G V P G W G I A L L V L 160
 510 540
 GTCTGTGTTCTGGTTGCGCTGGCCATTTGCTATCTCATTTGCTTTGGCTGTCTGTCAAGTGC
 V C V L V A L A I V Y L I A L A V C G C 180
 570 600
 CGCCGAAAGAACTACGGGAGCTGGACATCTTTCCAGCCCGGGATACCTACCATCCTATG
 R R K N Y G Q L D I F P A R D T Y H P M 200
 630 660
 AGCGAGTACCCACCTACCAACCCATGGGCGCTATGTGCCCTAGCAGTACCGATCGT
 S E Y P T Y H T H G R Y V P P S S T D R 220
 690 720
 AGCCCTATGAGAAGGTTTCTGCAGGTAATGTTGGCAGCAGCCTCTCTTACACAAACCA
 S P Y E K V S A G N G G S S L S Y T N P 240
 750
 GCAGTGGCAGCCACTTCTGCAACTTGTAG
 A V A A T S A N L U

Fig-6D

10/15

ATGACACCGGGACCCAGTCTCCTTTCTTCTGCTGCTGCTCCTCACAGTCTTACAGTT
 M T P G T G S P F F L L L L L T V L T V
 30 60
 GTTACAGGTTCTGGTCATGCAAGCTCTACCCAGGTGGAAGAAAAGGAGACTTCGGCTACC
 V T G S G H A S S T P G G E K E T S A T
 90 120
 CAGAGAAGTTCAGTGCCCAAGCTCTACTGAGAAAGAAATGCTCACTTCTCCCAAGTTGCTAC
 G R S S V P S S T E K N A H F S P V V Y
 150 180
 TGGGGTCTCTTTCTTTTTCTGTCTTTTACATTTCAAACCTCCAGTTTAA
 W G L F L F P V F S H F K P P V U
 210 60

Fig-7A

ATGACACCGGGACCCAGTCTCCTTTCTTCTGCTGCTGCTCCTCACAGTCTTACAGTT
 M T P G T G S P F F L L L L L L T V L T A
 30 60
 ACCCAGCCCTTAACCCGCAACAGTTGTTACAGGTTCTGGTCATGCAAGCTCTACCCCA
 T T A P K P A T V V T G S G H A S S T P
 90 120
 GGTGGAGAAAAGGAGACTTCGGCTACCCAGAGAAGTTCAAGTCCCAAGCTCTACTGAGAAG
 G G E K E T S A T G R S S V P S S T E K
 150 180
 AATGCTCACTTCTCCCAAGTTGCTACTGGGGTCTCTTTCTTTTTCTGTCTTTTACAT
 N A H F S P V V Y W G L F L F P V F S H
 210 240
 TTCAAACCTCCAGTTTAA
 F K P P V U

Fig-7B

11/15

ATGACACCGGGCACCCAGTCTCCTTTCT³⁰TCCTGCTGCTGCTCCTCACAGTGTCTTACAGTT⁶⁰
 M T P G T Q S P F F L L L L L T V L T V²⁰
 GTTACAGGTCTCTGGTCATGCAAGCTCTACCCAGGTGGAGAAAAGGAGAGCTTCGGGTACC¹²⁰
 V T G S G H A S S T P G G E K E T S A T⁴⁰
 CAGAGAAGTTCAAGTCCCCAGCTCTACTGAGAAGAATGCTATCCACGACCCGACTACTACC¹⁸⁰
 Q R S S V P S S T E K N A I P A P T T T⁶⁰
 AAGAGCTGCAGAGAGACATTTCTGAAATGTTTTTGCAGATTTATAACAAGGGGGTTTTTC²⁴⁰
 K S C R E T F L K C F C R F I N K G V F⁸⁰
 TGGGCCTCTCCAATATTAAGTTCAAGGCCAGGATCTGTGGTGGTACAATTGA²⁷⁰
 W A S P I L S S G G D L W W Y N U

Fig-8A

ATGACACCGGGCACCCAGTCTCCTTTCT³⁰TCCTGCTGCTGCTCCTCACAGTGTCTTACAGTT⁶⁰
 M T P G T Q S P F F L L L L L T V L T A²⁰
 ACCACAGCCCTTAACCCGCAACAGTTGTTACAGGTCTCTGGTCATGCAAGCTCTACCCCA¹²⁰
 T T A P K P A T V V T G S G H A S S T P⁴⁰
 GGTGGAGAAAAGGAGACTTCGGCTACCCAGAGAAGTTCAAGTCCCCAGCTCTACTGAGAAG¹⁸⁰
 G G E K E T S A T G R S S V P S S T E K⁶⁰
 AATGCTATCCCAGCACCGACTACTACCAAGAGCTGCAGAGAGACATTTCTGAAATGTTTT²⁴⁰
 N A I P A P T T T K S C R E T F L K C F⁸⁰
 TGCAGATTTATAACAAGGGGGTTTTCTGGGCCTCTCCAATATTAAGTTCAAGGCCAGGAT³⁰⁰
 C R F I N K G V F W A S P I L S S G G D¹⁰⁰
 CTGTGGTGGTACAATTGA²⁷⁰
 L W W Y N U

Fig-8B

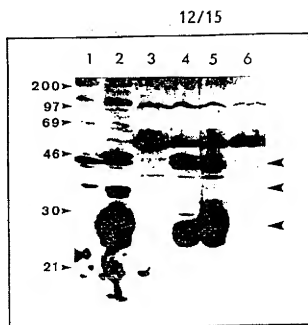


Fig-9a

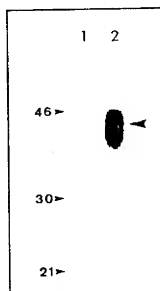


Fig-9b

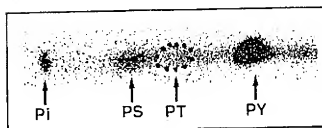


Fig-10

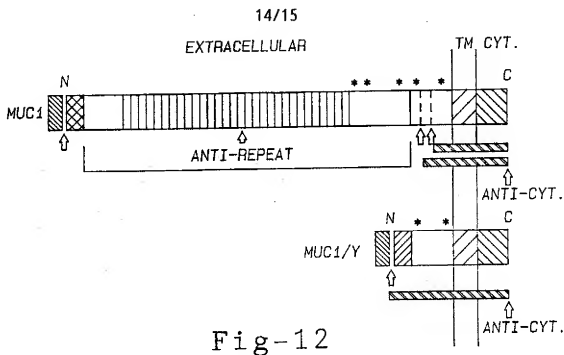


Fig-12

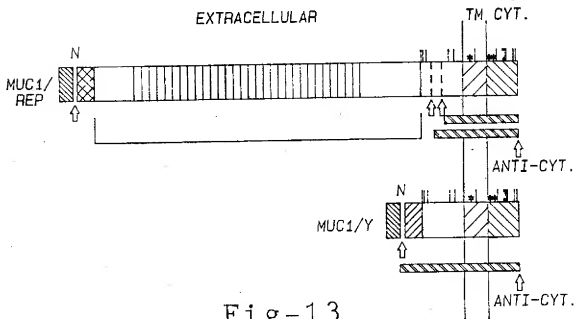


Fig-13

15/15

RECOGNITION SPECIFICITIES OF SH2 DOMAINS

<u>SH2 DOMAIN</u>	<u>pY+1</u>	<u>pY+2</u>	<u>pY+3</u>	<u>MUC1 SEQUENCE</u>
Src	E D T	E N T	I M L	YEKV (HUMAN) YEEV (MOUSE)
Fyn	E T	E D Q	I V M	
Lck	E D Q	E D Q	I V V	
Fgr	E Y D	E N D	I V	
Ab1	E T M	N E D	P V L	
Sem5/GRB-2	L V I	N P	P V	YTNP
PLC-gamma-1 C-TERMINAL	V I L	I L	P V I	YVPP

Fig-14